SCIENTIFIC AMERICAN books

SCIENTIFIC AMERICAN BOOKS are written and edited to open new frontiers of science to exploration by the general reader.

Each book is an illuminating account of one major area of research and is written by scientists intimately concerned with the work reported.

Each book derives from articles in SCIENTIFIC AMERICAN, the most authoritative—and the liveliest—magazine in its field ("Its roster of contributors," says *The New York Times*, "reads like a Who's Who of contemporary science").

The series has been greeted with widespread enthusiasm. The *Times* calls it "a stimulating source of information and reflection for the reader who is curious to learn about himself and the world he lives in." The *Bulletin of Atomic Scientists* writes: "They [SCIENTIFIC AMERICAN BOOKS] are of interest not only to the non-scientist but to the professional scientist who wishes to be *au courant* with the latest developments outside his specialization."

Simon and Schuster takes pleasure in publishing these lucid, authoritative books that present science as a vital human enterprise concerned with some of the largest questions that modern man has had the wit and courage to ask.

THE TITLES NOW AVAILABLE INCLUDE:

The New Astronomy
The Physics and Chemistry of Life
Twentieth-Century Bestiary
Atomic Power
Automatic Control
Plant Life
The Planet Earth
New Chemistry
The Universe
Lives in Science

EACH PRICED AT $1.45

THE UNIVERSE

How big is the universe? What is its shape? When did it start? Astronomy now reaches far enough into the depths of space and time to begin to answer the ancient questions of cosmology.

a SCIENTIFIC AMERICAN book

THE UNIVERSE

a
SCIENTIFIC AMERICAN
book

SIMON AND SCHUSTER · NEW YORK

PUBLISHED BY SIMON AND SCHUSTER, INC.
ROCKEFELLER CENTER, 630 FIFTH AVENUE
NEW YORK 20, N. Y.

FIRST PRINTING

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 57-7949
MANUFACTURED IN THE UNITED STATES OF AMERICA

TABLE OF CONTENTS

3. THE FORM OF THE UNIVERSE

Soon after the beginning, matter began to collect into great clumps which were the forerunners of today's galaxies. One group of stars in these island worlds was formed in that remote era. A second group is still in the process of being made. From the geometrical shapes of the galaxies we can read much of their history. When they collide, as they occasionally do, their interactions tell us a good deal about their internal constitution.

4. COSMOLOGICAL THEORY

There are at present two quite different views as to the large-scale structure of the universe. According to one, it originated with a big bang some five to six billion years ago and has been expanding ever since, growing steadily less dense as it fills more and more space. The other idea is that, although the universe is expanding, matter is being continuously created just fast enough to keep the average density constant. It should soon be possible to choose between the two models.

5. OBSERVATIONAL TEST

To discover the geometry of the universe, and learn whether it is evolving or in a steady state, requires observations of extremely distant galaxies. One crucial test is the amount by which their light is shifted toward the red end of the spectrum. Another is whether distant galaxies appear closer together than those nearby. This point is being investigated with both optical and radio telescopes.

6. CRITIQUE

Although they have an impressive body of observed fact on which to base their speculations, cosmologists must still make some rather large guesses. The author urges that these guesses be kept within the framework of accepted scientific method.

INTRODUCTION

No one who looks at present-day studies of the universe can fail to be impressed by the remarkably detailed foundation of theory and fact that has been established, nor by the soaring, imaginative structure erected on top of it. Consider some of the questions that astronomers and cosmologists are now asking: How big is the universe? What is its shape? When did it start? When, where and how was its matter manufactured? Was all its material present at the beginning or is it being continuously created? These are no longer mere expressions of childlike curiosity. They are meaningful queries to which science expects—and is getting—answers.

Many of the subtlest of current views will no doubt seem obvious to our children. Others will turn out to be wrong. Yet, even knowing that the present structure of ideas must inevitably fade into something different, one cannot but feel proud to be a member of the generation that is building it.

This book aims to display the whole edifice so that the uninitiated reader can admire its beauty and appreciate something of the cunning that has gone into its construction. The chapters appeared originally as articles in the September, 1956, issue of SCIENTIFIC AMERICAN, *devoted entirely to the subject of the universe. The issue provided a unique platform from which the distinguished authors could present, in straightforward, nontechnical language, a unified and up-to-date picture of their subject.*

In cosmology, as in every other field, today's ideas are simply the latest in a long, continuing struggle for understanding. Howard P. Robertson, in his introductory chapter, traces the his-

tory of man's attempts to discern in the world around him a coherent, comprehensible system. Babylonian shepherds were the spiritual ancestors of the men on Palomar Mountain. Greek geometers founded the line that produced Isaac Newton and Albert Einstein. Cosmology is a tapestry woven by astronomers, mathematicians and physicists.

One of the chief contributions of the physicist, William A. Fowler points out, is in deducing how the matter in the universe was built up from the primeval substance. The cosmic production sequence has now been almost completely figured out, and most of the processes can be duplicated and studied in the laboratory. The synthesis of heavy elements is apparently still going on in the interiors of stars.

Although matter may once have been spread out evenly through space, it has long since collected into the enormous clumps called galaxies. These are the units with which the cosmologist works. Walter Baade describes the present anatomy of galaxies, and Jan H. Oort tells how they appear to have arrived at their present condition. On the average, galaxies are separated from each other by millions of light years. However, some of them must occasionally bump into each other. One such colliding pair has recently been discovered and, as Rudolph Minkowski explains, it is yielding a great deal of exciting new information about the universe.

It is now generally agreed that the universe is expanding; that groups of galaxies are rushing away from each other with speeds which increase with the separation between them. There is, however, a sharp division of opinion as to the history of this expansion. One school holds that it started at a definite moment, some five or six billion years ago, with the explosion of a super-dense blob containing all the matter in the universe. The other view is that the expansion has always been going on at just its present rate, and that it will continue to go on forevermore with new matter being continually created to replace the old matter that is moving

away. The first, or evolutionary, theory is expounded by George Gamow. He discusses the various possible forms of an evolving universe. It may be closed and finite or open and infinite. It may have "begun" in an infinitely expanded state in the infinite past, contracted once to a minimum size and then started to expand again; alternatively, it may have contracted and expanded cyclically an indefinite number of times. The second view is put forward by Fred Hoyle, one of its leading exponents. He demonstrates the mathematical logic behind the rather startling steady-state theory, and explains how the question may soon be decided.

One of the pieces of evidence that will help settle the matter is the red shift, from which the fact of expansion was deduced. The more distant the galaxy, Allan R. Sandage explains, the farther its light is shifted toward the red end of the spectrum. This shift is interpreted as a Doppler effect, which means that the source of light is moving away from the earth. Still to be determined is the relation between distance and velocity for very remote galaxies. The light from these galaxies takes a billion years or more to reach us, and so we see them as they were a billion years ago. If their velocity is proportionately greater than for nearby galaxies, this will indicate that the expansion of the universe is slowing down, which would mean an evolving system rather than a steady state.

Another test depends on the density of galaxies. If they were closer together a billion years ago than they are now, the universe must be evolving. Jerzy Neyman and Elizabeth L. Scott describe how the comparative densities can be determined through statistical analysis. Martin Ryle tells us that radio signals will enable astronomers to extend their tests to much greater distances than those accessible to optical telescopes.

The modern cosmologist has a great deal of solid observational evidence to work with, but he is still forced to make some rather sweeping assumptions. In particular, he must believe that the section of the universe he can see is a fair sample of what lies

beyond his view. In the final chapter Herbert Dingle discusses some of the implications of this assumption in its various forms.

THE EDITORS *

THE UNIVERSE IN PHOTOGRAPHS

PLATE 1

GREAT NEBULA in Andromeda is a spiral galaxy of much the same size and type as our own. The spiral is viewed at a shallow angle to its central plane. Above and below the galaxy in this 48-inch Schmidt-telescope photograph are two much smaller satellite galaxies of the elliptical type (see chapter on the evolution of galaxies, page 43).

PLATES 2–5

SPIRAL GALAXIES are shown more or less face on in these four photographs. Plate 2 is a photograph of NGC 7217, a galaxy of Type Sa: spirals with tightly wound arms. Plate 3 shows NGC 3031, a galaxy of Type Sb: spirals with less tightly wound arms. Plate 4 shows NGC 5194, a galaxy of Type Sc: spirals with still less tightly wound arms. The blob at lower left in this picture is actually NGC 5195, a companion of the larger galaxy. Plate 5 shows NGC 598, another galaxy of Type Sc with arms more open. The NGC number identifies the galaxy in the New General Catalogue, adopted in 1887 as the international registry for star clusters and extragalactic objects.

PLATES 6–7

SPIRAL GALAXIES, edge on in these photographs, reveal their characteristic thin cross section and central disk of dust and gas. This obscuring material is responsible for the dark lane against the bright background of stars. Plate 6 shows NGC 4565, a galaxy of Type Sb (see Plate 3). Plate 7 shows NGC

891, a galaxy of Type Sc. In the latter type the disk is less regular. The photographs in Plates 2–7 were made with the 200-inch telescope.

PLATES 8–9

BARRED-SPIRAL GALAXIES were also photographed with the 200-inch telescope. Plate 8 shows NGC 1398, a galaxy of Type SBa: a barred spiral with tightly wound arms. Plate 9 shows NGC 1300, a galaxy of Type SBb: a barred spiral with less tightly wound arms.

PLATE 10

ELLIPTICAL GALAXY NGC 4621 was photographed with the 200-inch telescope. Elliptical galaxies are characterized by their symmetrical mass of stars and lack of dust and gas.

PLATE 11

IRREGULAR GALAXY is represented by the Large Cloud of Magellan. This galaxy is one of the two nearest our own; the other is the Small Cloud of Magellan. The photograph was made with a wide-field 10-inch telescope at the Boyden Station in South Africa.

PLATES 12–13

STELLAR POPULATIONS are contrasted in these two photographs of the spiral galaxy NGC 4594, which we see edge on. Plate 12, made with a red filter, suppresses the bright blue stars in the disk of the galaxy and brings out the red Population II stars which are concentrated around and in the roughly spherical nucleus. Plate 13, made with a blue filter, reduces the brightness of the nucleus and brings out the light of the blue stars formed in the dust and gas that obscure the edge of the disk.

PLATE 14

GREAT GLOBULAR CLUSTER Omega Centauri was photographed with the 33-inch Baker-Schmidt telescope at the

Boyden Station. Globular clusters contain Population II stars and are spherically distributed around center of galaxy.

PLATE 15

POPULATION II STARS IN OUR GALAXY are shown in this 48-inch Schmidt photograph of the Milky Way in Sagittarius. The picture indicates the density of stars in the central plane of our galaxy and suggests the difficulty of resolving them in other galaxies.

PLATE 16

POPULATION II STARS IN ANDROMEDA NEBULA are resolved in this 200-inch telescope photograph. The brighter round obects and those surrounded by four spikes are stars within our own galaxy. Nucleus of the Andromeda Nebula is toward the left.

PLATE 17

RED SHIFT of four galaxies on Plate 17 is depicted in their spectra. The galaxies are centered in the photographs on the left-hand page. The spectra are the bright horizontal streaks tapered to the left and right in the spectrographs on the right-hand page. Above and below each spectrum are comparison lines from the spectrum of iron. Near the left end of the topmost spectrum are two dark vertical lines: the K and H lines of calcium. If the galaxy did not exhibit the red shift, these lines would be in the position of the broken line running vertically down the page. The amount of their shift toward the red, or right, end of the spectrum is indicated by the short arrow to the right of the broken line. The larger shift of the K and H lines of the three fainter galaxies is indicated by the longer arrows below their spectra. The constellation, approximate distance and velocity of recession of each galaxy is at left of its photograph.

PLATE 18

STRONGEST RADIO SOURCE, located in Cassiopeia, has been

identified with the faint nebulosities, apparently thin wisps of interstellar matter, shown in this 200-inch photograph.

PLATE 19

GALACTIC COLLISION in Cygnus is represented by the irregular spot in the center of this photograph made with the 200-inch telescope. Although the galaxies are 270 million light-years away, they comprise the second strongest radio source in the sky. This contrasty negative print shows only the distorted central masses of the galaxies. Their total visible diameter is about six times greater.

PLATE 20

ANOTHER PAIR OF COLLIDING GALAXIES is NGC 4038 and 4039. This negative print from a photograph made with the 200-inch telescope shows long tidal filaments above and below the central masses of the galaxies. The radio emission of this system is weak.

PLATE 21

NEAR-COLLISION of two spiral galaxies is also photographed with the 200-inch telescope. The galaxies are NGC 5426 and 5427. Tidal interactions between the two are faintly visible in the photograph. This system is not, however, a collisional radio source.

PLATE 22

A huge cluster of galaxies in the constellation of Coma Berenices is shown in this negative print. The photograph was made with the 48-inch Schmidt telescope on Palomar Mountain. The galaxies are the small fuzzy objects. They may be further distinguished from nearby stars by the fact that the images of the stars are either small and round or larger and surrounded by four spikes. The spikes are caused by the diffraction of starlight around the four structural members which support the photographic plateholder within the tube of the Schmidt telescope.

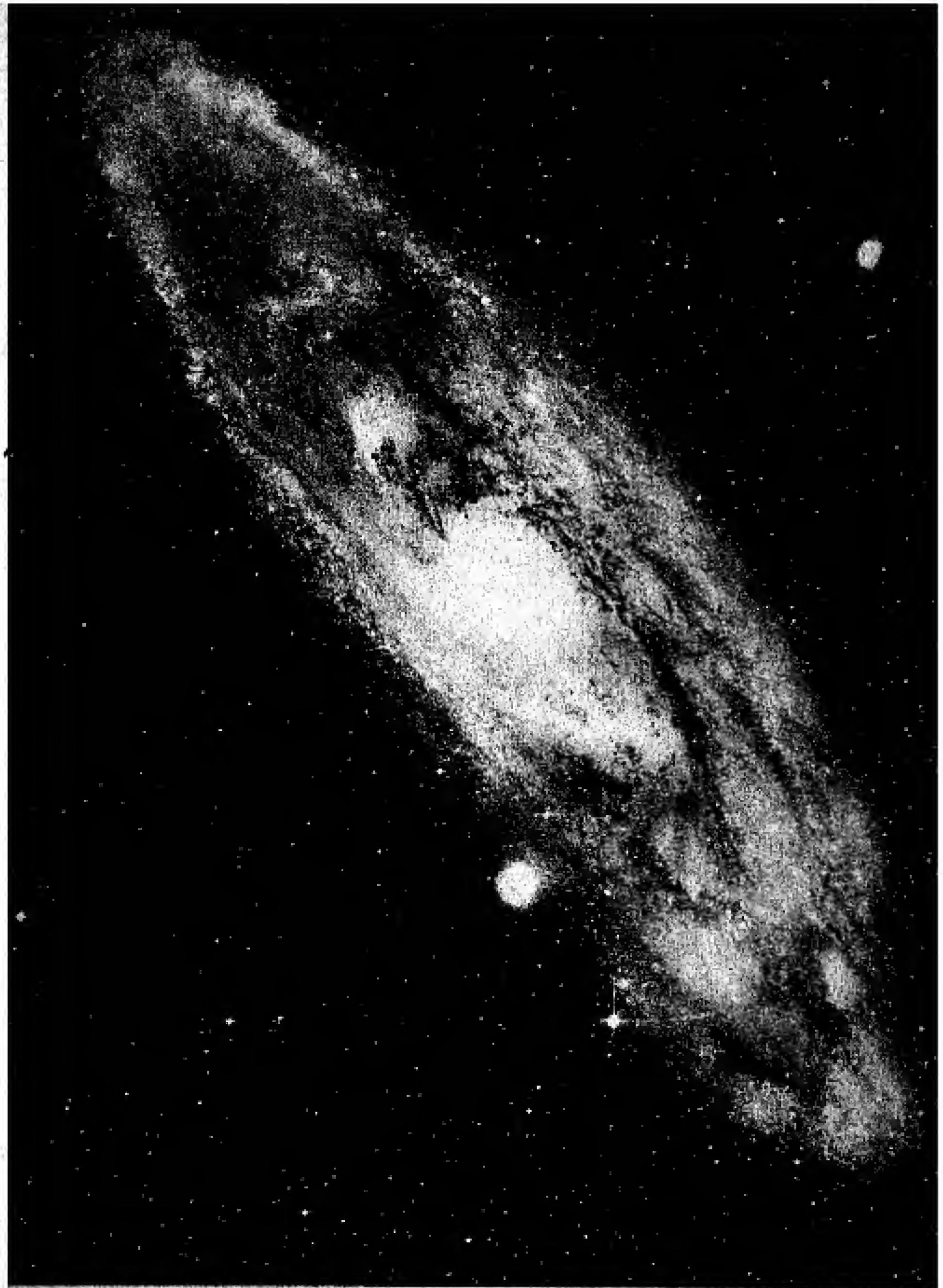


PLATE 1 GREAT NEBULA IN ANDROMEDA

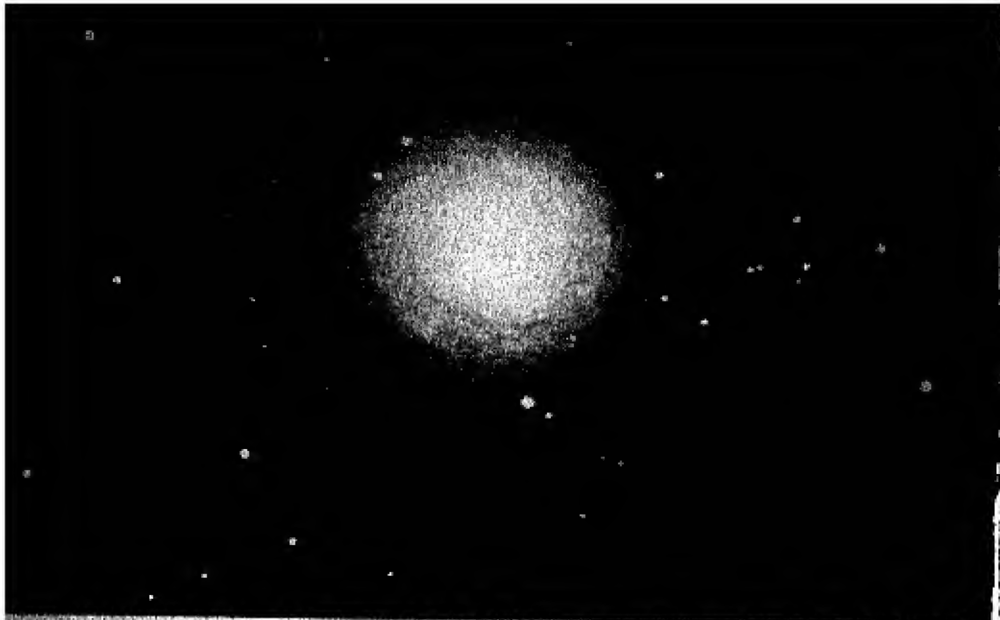


PLATE 2 SPIRAL GALAXY NGC 7217

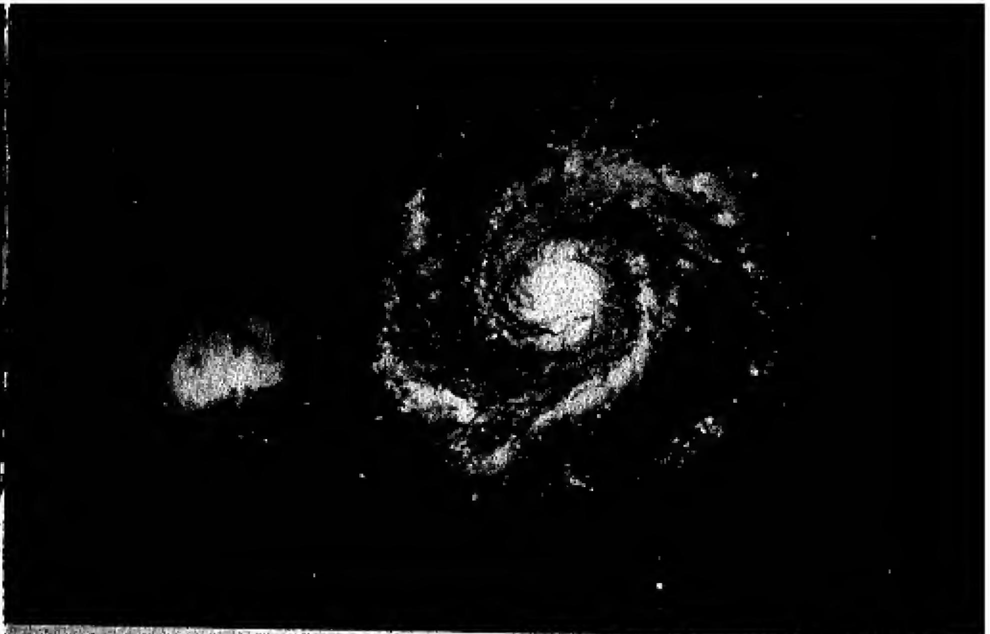


PLATE 4 SPIRAL GALAXY NGC 5194

PLATE 3 SPIRAL GALAXY NGC 3031

PLATE 5 SPIRAL GALAXY NGC 598

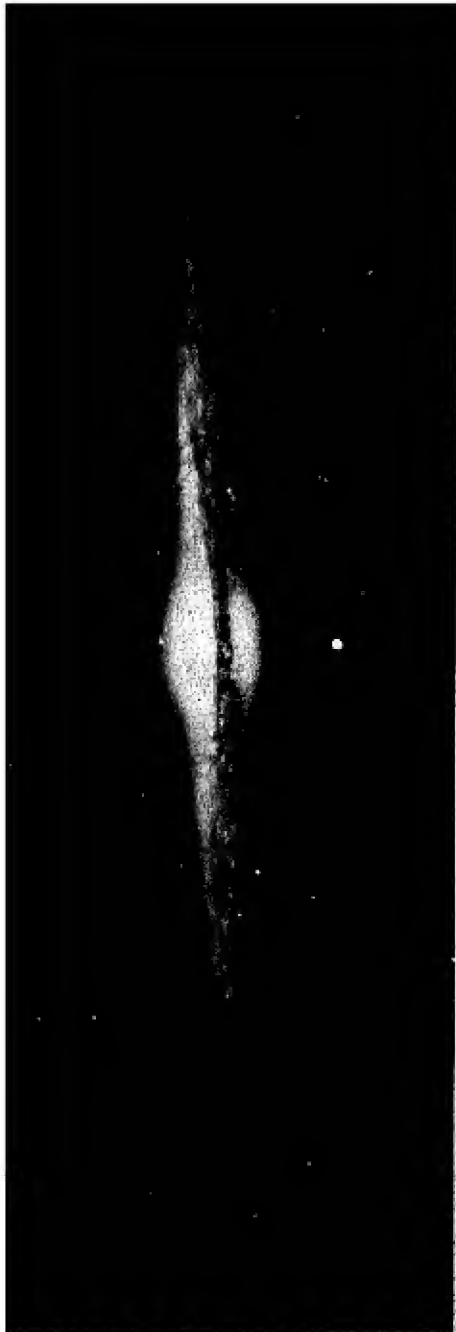


PLATE 6 SPIRAL GALAXY NGC 4565

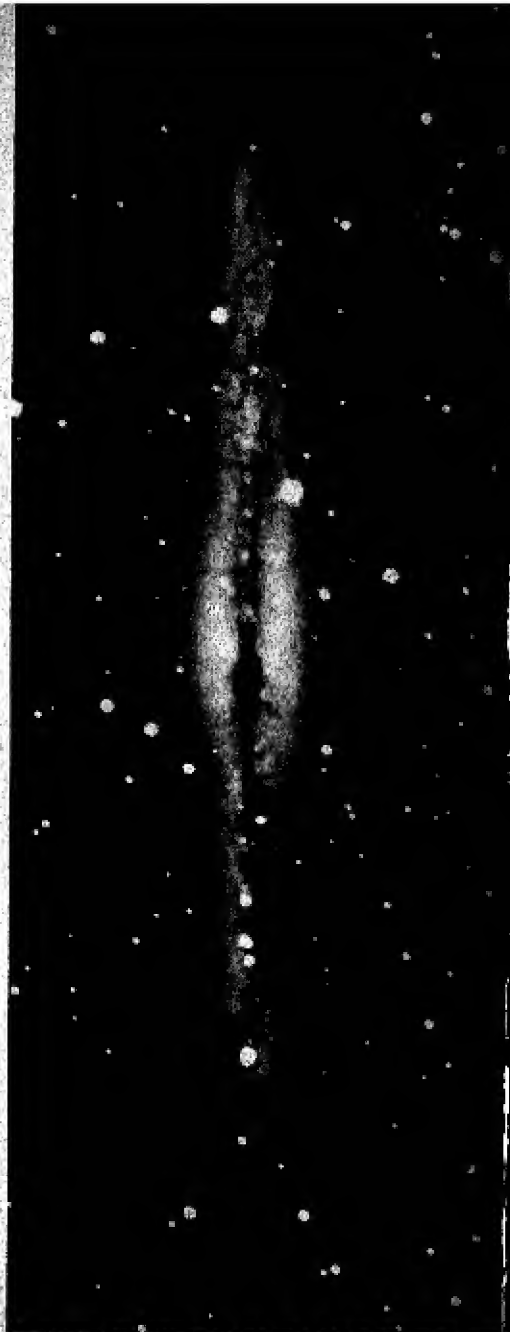


PLATE 7 SPIRAL GALAXY NGC 891

PLATE 8 BARRED-SPIRAL GALAXY NGC 1398

PLATE 9 BARRED-SPIRAL GALAXY NGC 1300

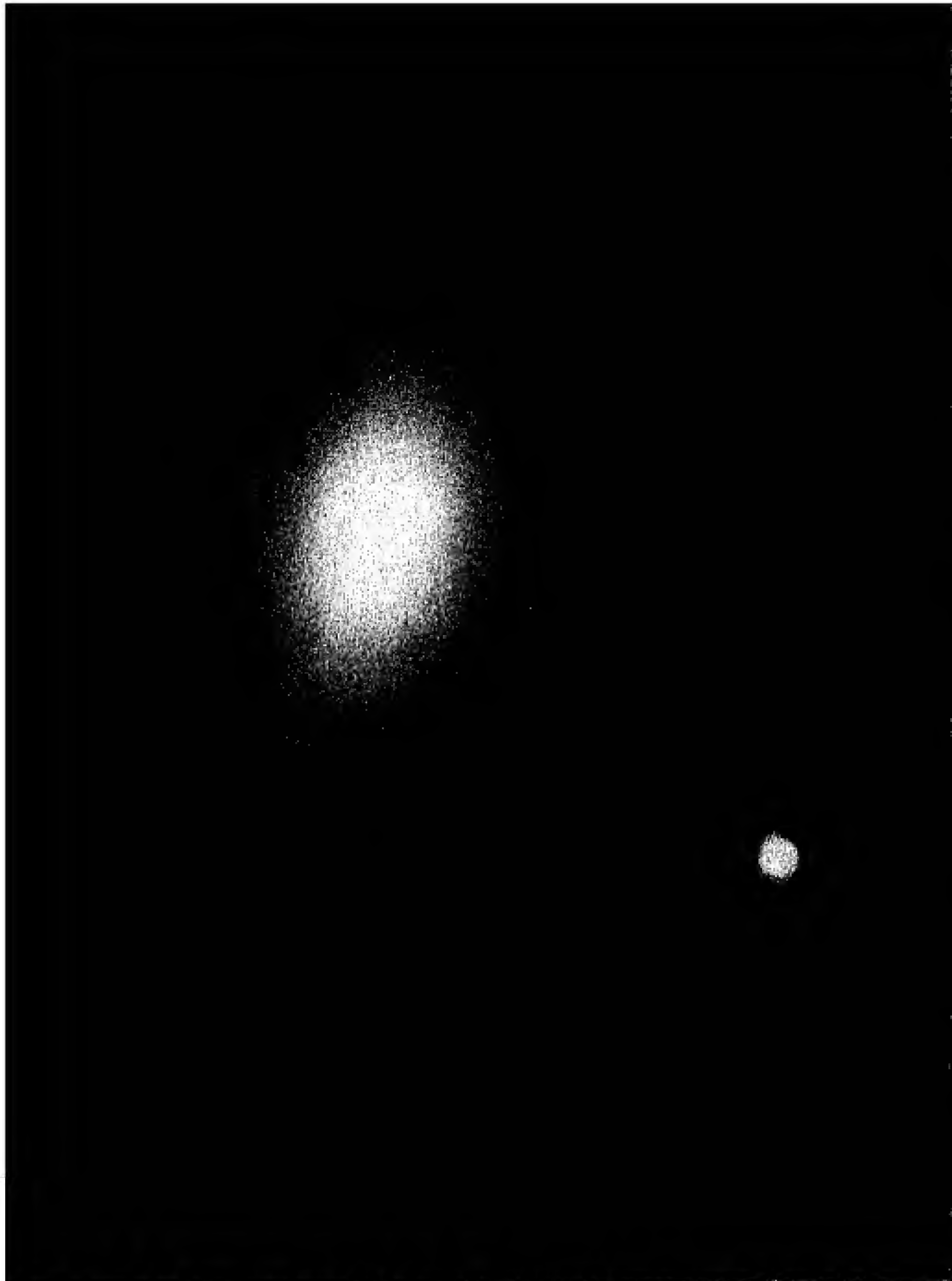


PLATE 10 ELLIPTICAL GALAXY NGC 4621

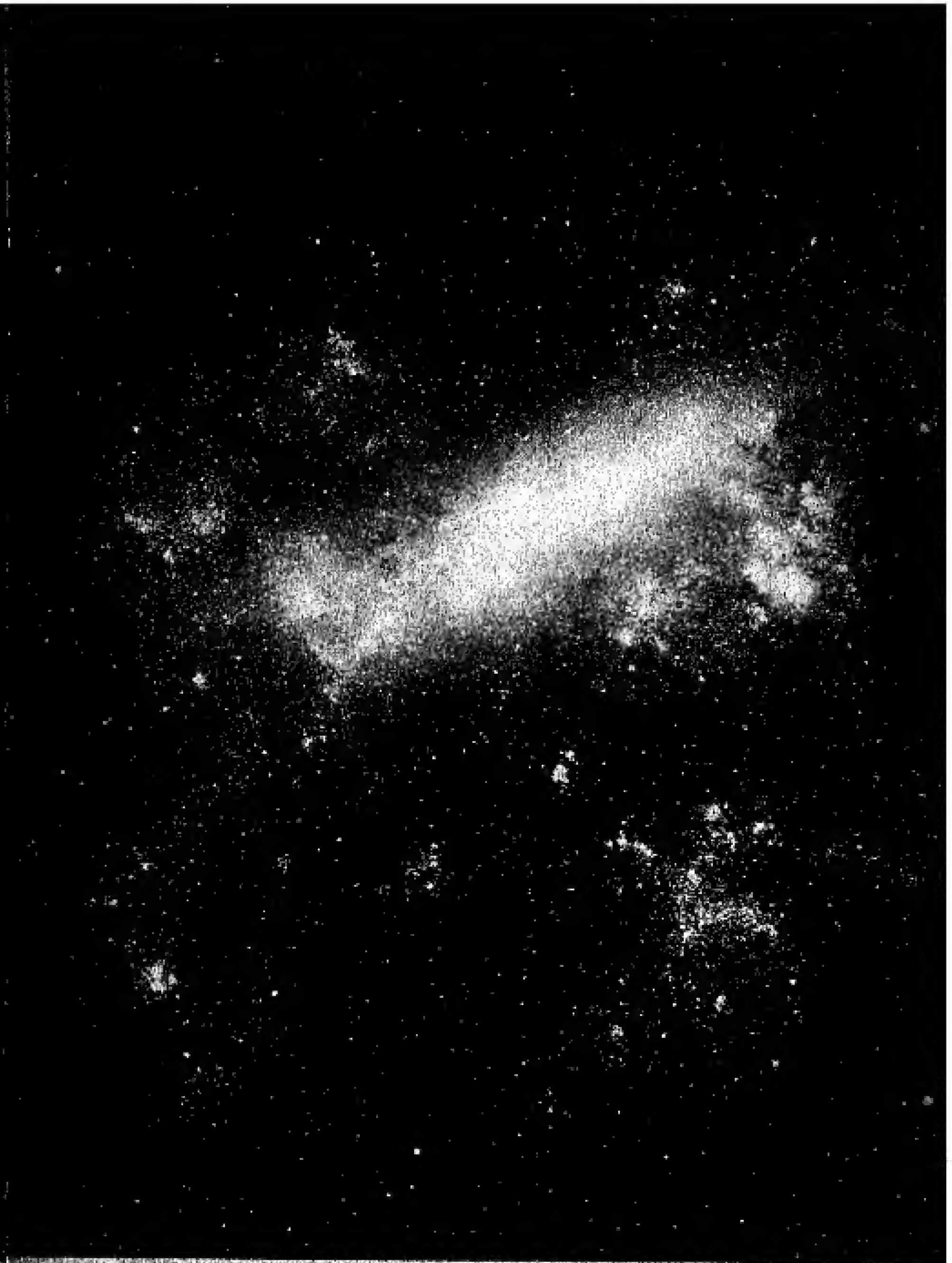


PLATE 11 IRREGULAR GALAXY MAGELLANIC CLOUD

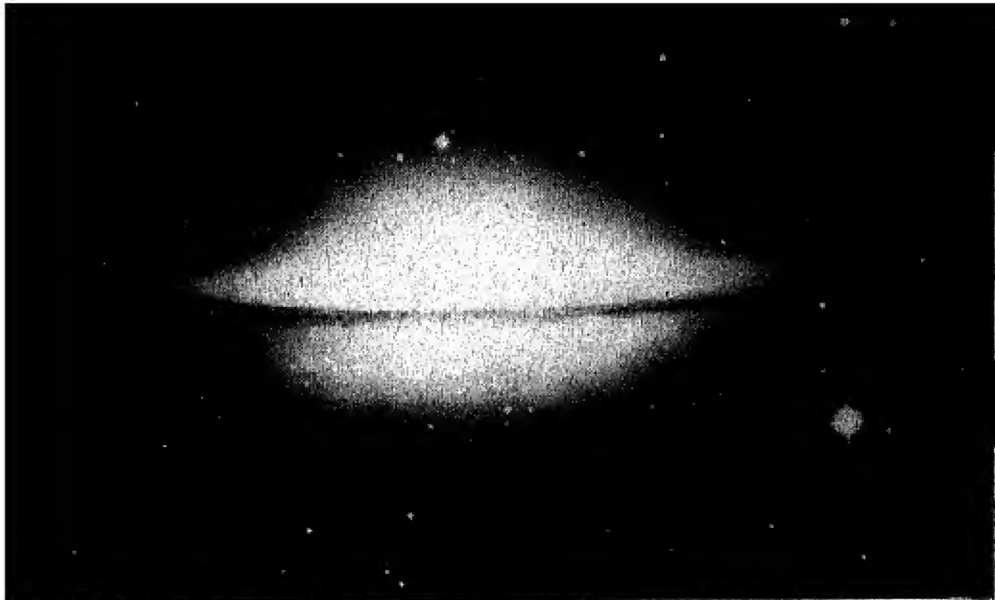


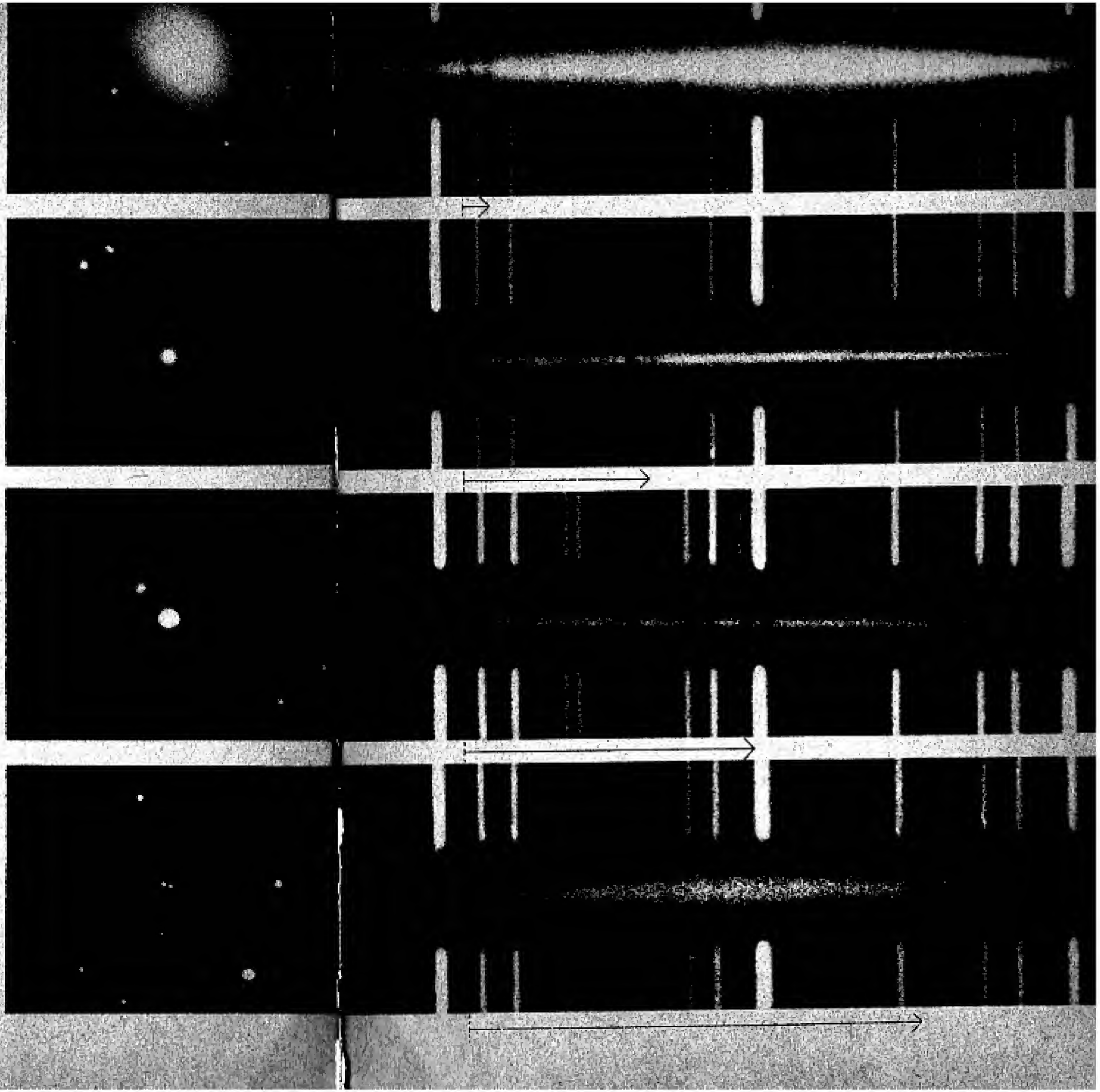
PLATE 12 STELLAR POPULATIONS IN THE GALAXY NGC 4594—RED FILTER

PLATE 13 STELLAR POPULATIONS IN THE GALAXY NGC 4594—BLUE FILTER

PLATE 14 GREAT GLOBULAR CLUSTER OMEGA CENTAURI

PLATE 15 POPULATION II STARS IN OUR GALAXY

PLATE 16 POPULATION II STARS IN ANDROMEDA NEBULA



VIRGO
22 Million Light-Years
1,200 Kilometers Per Second

CORONA BOREALIS
400 Million Light-Years
21,500 Kilometers Per Second

BOOTES
700 Million Light-Years
39,300 Kilometers Per Second

HYDRA
1.1 Billion Light-Years
60,900 Kilometers Per Second

PLATE 17 RED SHIFT OF FOUR GALAXIES

PLATE 18 STRONGEST RADIO SOURCE

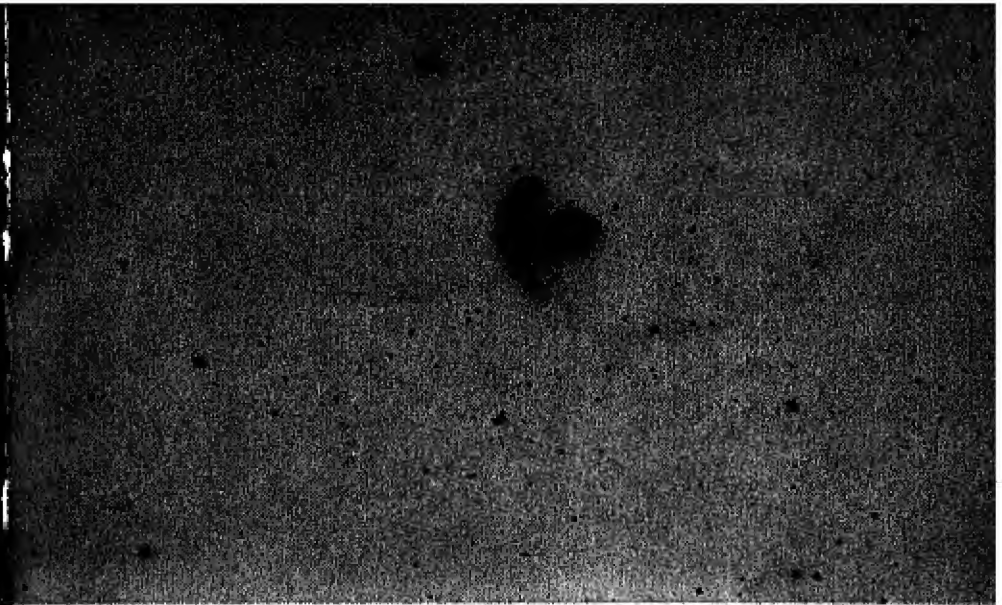


PLATE 20 ANOTHER PAIR OF COLLIDING GALAXIES

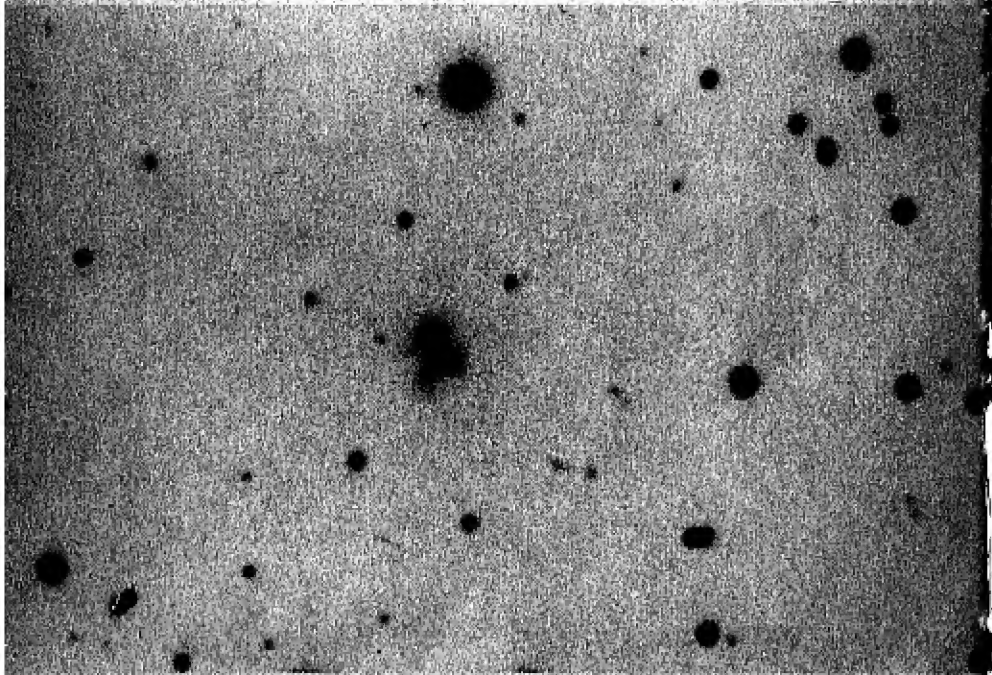


PLATE 19 GALACTIC COLLISION IN CYGNUS

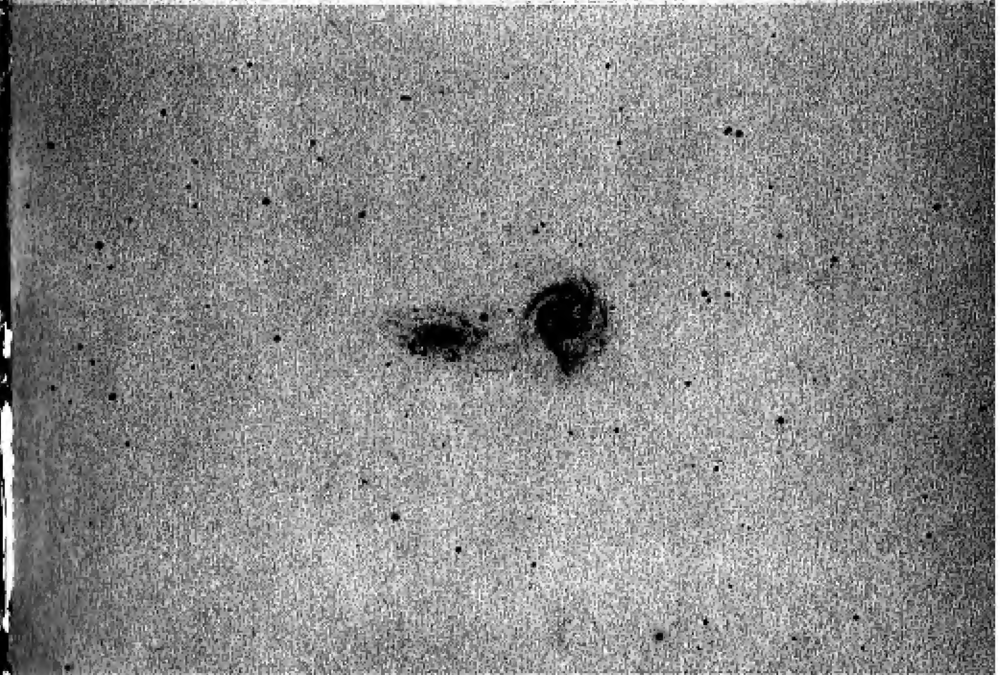


PLATE 21 NEAR-COLLISION OF TWO SPIRAL GALAXIES

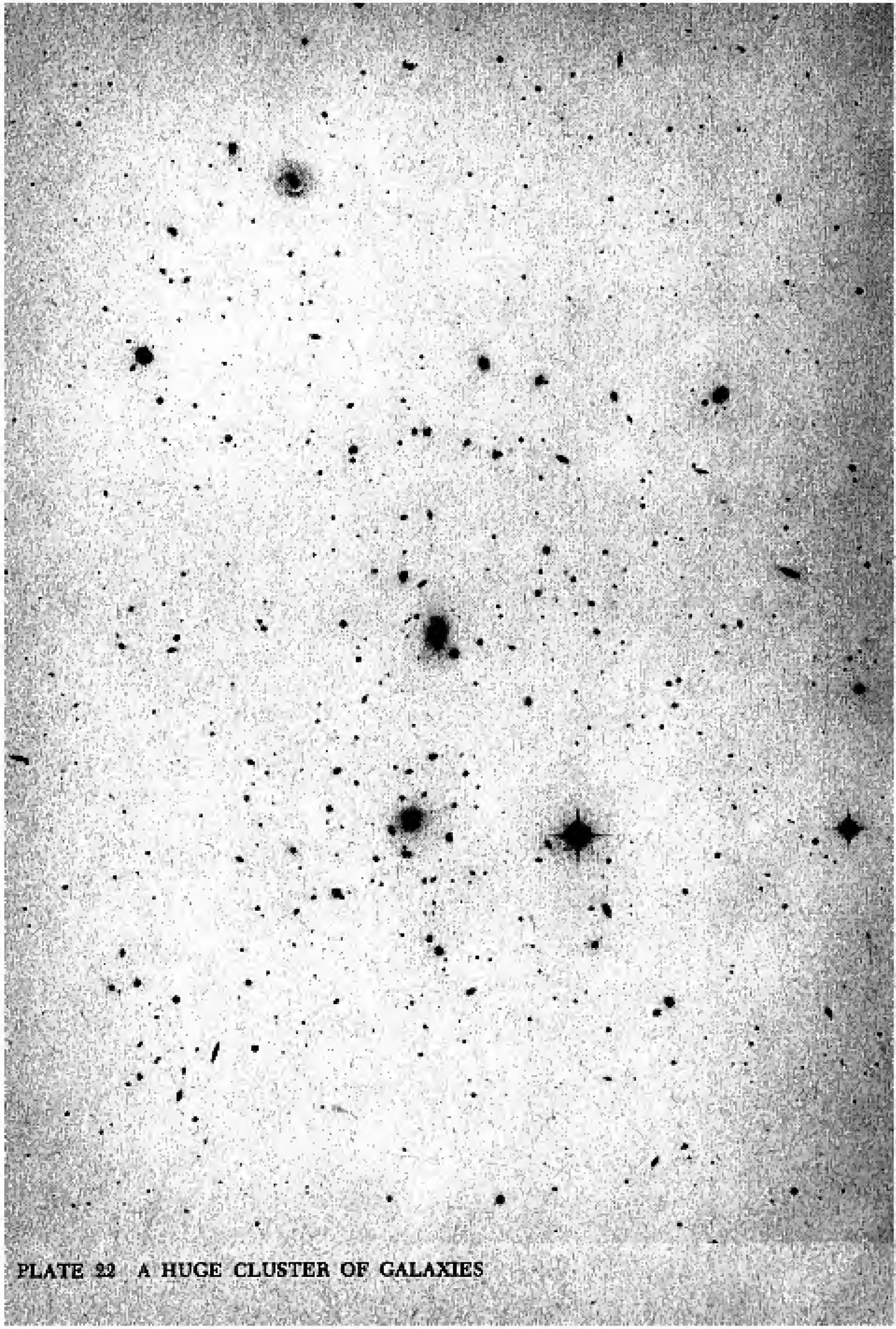


PLATE 22 A HUGE CLUSTER OF GALAXIES

PART 1 HISTORY

THE UNIVERSE *by Howard P. Robertson*

The author is professor of mathematical physics at the California Institute of Technology. After acquiring his doctorate at Cal Tech in 1925, he studied for three years under a National Research Council fellowship in Göttingen, Munich and Princeton. He was a member of the Princeton University faculty from 1928 to 1947, when he returned to Cal Tech. He first became interested in cosmology while studying in Europe. At that time he found a solution of Einstein's gravitational equation which seemed to explain the recession of nebulae that had been noted by V. M. Slipher of the Lowell Observatory. Investigating this and other applications of relativity theory to astronomy occupied much of his attention for more than a decade. He revised a prerelativistic suggestion of J. H. Poynting: that a particle absorbing and re-emitting radiation from the sun would eventually spiral into the sun. Now known as the Poynting-Robertson effect (says Robertson: "*I* didn't name it!"), this idea has proved important to meteor astronomy and cosmogony. The outbreak of World War II brought Robertson back to earth and classical mechanics. He did theoretical studies on the effects of bombs and projectiles, and later got into operations research. After the war he was again called to Washington to direct research for the Weapons Systems Evaluation Group and, in 1954, he was appointed scientific adviser to the Supreme Allied Commander in Europe. Finally, in 1956, he was able to return to Cal Tech and a full-time academic career.

THE UNIVERSE

by Howard P. Robertson

LOOKING OUT on the night sky we are deeply impressed, as the billions of human inhabitants of this planet have been before us, by the majestic procession that wheels unceasingly by. Even early man saw, and speculated about, both the regularity and the irregularities of this nightly scene: the myriad stars rolling steadily across the heavens, the slight changes in the vista from one night to the next, the moon slowly wending her way backward through the star-studded firmament, the planets ("wanderers") roaming erratically through the field, the occasional fall of a "shooting star" and, less frequently, the eerie sweep of a comet.

Quite naturally early man tended to think of himself as the center of it all—the hub about which this great spectacle turned. He ordered these celestial appearances, like the immediate physical surroundings with which he had to grapple, into a self-centered world—and sat serenely in the middle, the favored onlooker if not the master of all he surveyed.

What happened to jut man out of this complacent rut? By what turn of the intellect did he come to relegate himself to his present position as the denizen of a middle-of-the-road satellite of an undistinguished star in a galaxy which itself is but an ordinary member of an uncountable universe of galaxies? How did this cosmic view develop, and what sort of universe do we see today?

These are the questions that the authors of the succeeding chapters in this book seek to answer, insofar as the present state of knowledge will allow. They tell of the current investigations, both observational and theoretical, into the make-up and history of galaxies and the age and extent of the visible universe as a

whole. In this introductory account I shall sketch with a broad brush the path that man has taken in exploring his relation to the cosmos, dwelling briefly upon the major turning points.

Modern scientific cosmology is descended from a long line of thought which arose early in the history of civilization in Babylonia and in Egypt, was developed by the Greeks, was preserved and elaborated by the Arabs through the Middle Ages, was reconstructed by scholars of the Renaissance and was thus handed down to the men who transformed it and thereby inaugurated the present age of science. The scheme built by the Greeks and their followers was founded upon (1) an urge to understand the celestial system, and (2) a willingness to settle for a purely descriptive account of its motions. Their model of the universe grew out of Aristotelian physical principles: namely, that the earth and everything else under the moon was made up of four elements—earth, water, air and fire—which tended to move along the vertical to regain their natural positions in a series of spherical shells concentric with the earth, while the celestial bodies were formed of the quintessential ether, and in keeping with their unchanging perfection must maintain their position above the earth. The celestial bodies were thereby restricted to motion in concentric orbits and must be firmly attached to tangible spheres, because according to Aristotle a body can remain in motion only when in physical contact with something that causes it to move. Thus grew up the elaborate system of homocentric spheres carrying the moon, sun, planets and stars. Behind the sphere of fixed stars was presumably a Prime Mover, responsible for all the complicated motions of the spheres. Beyond the stars was the void: the finite, spherical universe was surrounded by nothing, not even space.

Ingenious as this system was, it was simply not capable of accounting for some of the motions and other phenomena actually observed. To "save the phenomena" the later Greeks fell back upon a purely mathematical representation of motions intended to

account for observed irregularities. Its most characteristic device was to allow each of the seven wanderers (including the sun and moon) to move uniformly in a circle (the epicycle) whose center itself moved uniformly on another cycle (the deferent) about the earth. Since the whole purpose of the exercise was to account for the motions as viewed from the earth, this was quite naturally taken as at rest, and the firmament of stars allowed to rotate daily about it. In the form elaborated by Ptolemy in the second century, this geometrical scheme was capable of giving a tolerably accurate account of the naked-eye observations amassed over the preceding centuries. It was, however, truly a scheme rather than a system, for it embodied no physical principle accounting for the relations between the orbits of the various bodies. Ptolemy's value for the mean distance of the moon from the earth was accurate to within 2 per cent, but his estimate of the distance of the sun was only a twentieth of the true value. The root of the trouble lay in the fact that Greek physics was woefully inadequate to the tasks put upon it.

The extrication of cosmology from this welter of philosophy and astronomy began when Nicolaus Copernicus early in the sixteenth century transferred the central reference point from the earth to the sun. Copernicus retained the epicyclic motions to save the phenomena, but by allowing the earth to rotate he stopped the dizzy motion of the stars, and by allowing the earth to revolve about the sun he did away with the complications Ptolemy had had to introduce in the planetary motions to offset the assumed motion of the sun around the earth. The sphere of the fixed stars was now at rest, and taken to be large enough so that the earth's pilgrimage about the sun would result in no noticeable displacement of the stars in the sky from opposite points in its orbit. (It was not until the nineteenth century that meticulous measurements made it possible to observe such parallactic displacements for nearby stars.)

Toward the end of the sixteenth century Thomas Digges, the first to present the essentials of Copernicus' *De Revolutionibus*

Orbium Coelestium in English, introduced a change in the Copernican scheme which at first sight seemed so slight that even Digges failed to appreciate its basic significance. In his *Perfit Description of the Coelestiall Orbes,* Digges replaced the sphere of fixed stars with an infinity of stars extending uniformly throughout an infinite universe. This change made inevitable the eventual dethronement of the sun, which, in Digges's own words, "like a king in the middest of al raigneth and geeveth lawes of motion to ye rest." But the time was not yet ripe for such a reorientation: its full fruition would require the inventive genius and skill of the great observers from Galileo Galilei to George Ellery Hale, and the unifying genius and thought of the host of theorists from Isaac Newton to Albert Einstein.

Three of the elements needed for the breakthrough to the Newtonian conception of the world were furnished around the turn of the seventeenth century by Galileo and by Johannes Kepler. Galileo's service was twofold, both in the observational domain and in furthering the breach with the impotent Aristotelian dynamics by recognizing clearly that only change of motion —and not motion itself—was impelled by the force applied. His development of the telescope into a practical instrument enabled him, by discovering the satellites of Jupiter, to strip the earth forever of its claim to being the only center of rotation in the universe. At the same time he resolved portions of the Milky Way into great collections of stars, paving the way for the dethronement of the sun from the central position assigned it by Copernicus. Curiously enough Galileo held to the doomed epicycles. It was Kepler who, in order to save the phenomena as observed by Tycho Brahe, the last great naked-eye astronomer, threw off at last the restricting dogma of uniform motion in circles. His three laws describing the motions of the planets, traveling in elliptical orbits, supplied the cornerstone for a rational structure of the solar system, though Kepler himself failed in his lifelong attempt to put the pieces together.

What was needed was a unifying principle to bring order into the disjointed parts, to tie them together into a physical model which would satisfy the philosophers in their quest for truth and the astronomers in their requirement of accuracy. This Newton supplied in his theory of universal gravitation. He subjected not only the solar system but also the ethereal stars to one reign of law. Gravitation was the force that held objects on the earth, that bound the moon to it, that steered the planets in their elliptical orbits about the sun and adjusted their periodic times in accordance with their distances from it. For the first time it was possible to measure accurately the dimensions of the solar system. Along with his gift of gravitation, Newton invented the tools required for the construction of a universal system by completing the science of dynamics and perfecting the infinitesimal calculus.

The first crude steps could now be taken toward a cosmology describing the world at large, and toward a cosmogony accounting for the evolution of its parts. Newton himself speculated that "if the matter were evenly distributed throughout an infinite space . . . some of it would convene into one mass and some into another, so as to make an infinite number of great masses, scattered great distances from one another throughout all that infinite space. And thus might the sun and fixed stars be formed. . . ." This bold leap into the distant reaches of time and space blazed a path for Immanuel Kant, Pierre Simon de Laplace and other cosmogonists to follow. But the vastness and complexity of the system that would have to be explained was not yet even suspected: there was still little idea of the organization of stars in loose associations and tight clusters or of the various types of nebulae, to say nothing of the clustering of tens of thousands of these island universes into supersystems so vast that it takes light tens of millions of years to traverse them.

By the first quarter of the present century the ever-quickening accumulation of research had set the stage for spectacular developments. The telescope, increasing in power from the early re-

flectors of Sir William Herschel and of the Earl of Rosse to the great instruments of America, had shown the Milky Way system to be a lens-shaped collection of some 100 billion stars, with the sun out toward the rim. The spectroscope, analyzing the light of the stars, had identified family traits among them. Studies of their color shifts, assumed to be a Doppler effect, had told something of the stars' motions away from us. The distances of stars had been measured by the methods of parallactic displacement and comparison of apparent brightness. Henrietta Leavitt at the Harvard College Observatory and Harlow Shapley at Mount Wilson had discovered that the pulsating stars called Cepheid variables provided a scale for measuring distances in our own galaxy, the Milky Way, and for estimating the distances of other galaxies in which these variables could be resolved. With this new tool it was found that the sun's distance from the center of our galaxy was some 26,000 light-years—a far cry indeed from the little world of Johannes Kepler, which light could have traversed in less than one tenth of a year! And the galaxy was found to be spinning about that distant center at a rate which would swing the sun around it in 200 million years.

The galaxy itself had been measured, and many of the objects seen in the heavens had been identified as members of it. But what of the multitude of nebulae, now numbering thousands, that could not be placed in our system? Immanuel Kant had in 1755 put forward the notion that such nebulae were themselves distant island universes—peers in every way of our own galaxy. The proof had to await the development of methods of measuring the titanic distances involved. The break came in 1917, when G. W. Ritchey at Mount Wilson identified a nova in the Great Spiral Nebula in Andromeda (see Plate 1). By 1924 Cepheid variable stars had been found in this and other nearby nebulae, and on the heels of this came the development of other brightness criteria of distance—supergiant stars, supernovae, nebulae themselves and finally even the brightest nebulae in clusters. This break initiated the modern era in cosmology.

One of the first tasks was to identify the galaxies by type. Edwin P. Hubble at Mount Wilson established a system which starts with elliptical nebulae, of increasing degrees of flatness, and branches off into normal spirals on the one hand and barred spirals on the other, with the sprinkling of irregular nebulae falling outside the sequence. The question immediately arose: Where does our own galaxy fit into the scheme? The unevenness of its texture, its great star clouds and murky patches, indicate that it does not belong in the class of elliptical nebulae, which are almost structureless. On the other hand, its rotation and rudimentary symmetry suggest that our galaxy is not to be classed with the irregular nebulae, such as our closest companions, the Clouds of Magellan (see Plate 11). Hubble tentatively placed the Milky Way well toward the end of the branch of normal spirals, a counterpart of the very open spiral in Triangulum (see Plate 5). It is indeed easy to imagine that an observer looking out at that galaxy from a point near its rim would see a ragged veil of stars crossing his sky as does our own Milky Way; its nucleus would be obscured, like ours, by dark clouds of dust, and the structure of its arms would be unclear, again like ours, because the observer would be looking at them along the plane in which they lie.

But science will not rest with satisfying the imagination; the imagined picture is a challenge driving the astronomer on to find means of mapping the elusive nucleus and the spiral arms, if such there be. And such means are being found, making the problem of galactic structure one of the most active and exciting in the whole cosmological field. Closer study of the spiral in Andromeda has brought out certain characteristics of spirals—red giant stars and cluster-type variables in the nucleus, blue supergiants surrounded by hydrogen gas in the arms. These features serve as clues for exploring our own galaxy. A marvelous new tool, radio astronomy, has enabled us to detect hydrogen gas lining its spiral arms, and to hear significant radio "noises" from its central regions. The evidence indicates more and more that we are resi-

dents in a normal spiral galaxy, very similar in size, structure and composition to the Andromeda Nebula.

Having found a way to identify and classify galaxies, the next task was to map their distribution. Surveys showed that nebulae tend to cluster in groups, containing up to a thousand or more. The exploration has in fact suggested to some that our own galaxy may be an outrider in a supergalaxy, just as the sun was earlier found to be an outrider of the galaxy.

Are the clusters themselves distributed irregularly or are they scattered at random more or less uniformly through space? The evidence is conflicting, but inspection of selected regions indicates that by and large the fainter—and therefore presumably more distant—nebulae are indeed fairly uniformly distributed, aside from their apparent tendency to cluster. Here again is a field of research which is being explored vigorously, both from observational and theoretical points of view.

The question of the distribution of nebulae in depth—that is, with increasing distance from us—is a crucial one for deducing the structure of the universe as a whole. According to our recently recalibrated distance scale the 200-inch telescope on Palomar Mountain reaches out to more than two billion light-years, and new radio telescopes may take us even farther. If the concentration of nebulae is found to change with distance, we may obtain a measure of the curvature of space postulated by certain interpretations of Einstein's general theory of relativity. If nebulae at great distances are found to differ physically from those nearby, we may have clues to the evolution of galaxies, for we are seeing the distant galaxies as they were billions of years ago, when their light began its enormous journey to us.

These possibilities raise the question of the time scale of the universe. Did the universe originate at some finite time in the past, or has it existed forever? The question may not be one for science alone to answer, but there can be no doubt that it is one to which the findings of science are relevant. Many of the older

cosmologies assumed that the universe was created full-blown in very much the state in which we find it now; more recent theories, in rationalistic reaction, tacitly assumed that the universe had no beginning. If the first of these assumptions were correct, it would be hard to account for the observed existence of one-way processes in nature; the second fails to account for the continued existence of radioactivity. The nebulae themselves, by their motions, offer an answer to this question of time—if not a complete answer, then at least a time scale which must prove significant for any cosmological theory. For the nebulae appear to be rushing away from us; if these motions are real, and if they have been going on continuously in the past, then all the matter we now see spread throughout the universe must a few billion years ago have been compressed into a very small compass indeed. It is as if, at that time in the dim past, there was a mighty explosion which catapulted fragments out to form the nebular universe as we now see it.

The evidence for this view is, of course, the well-known red shift in the light we receive from distant nebulae. By the end of the first quarter of this century V. M. Slipher at the Lowell Observatory at Flagstaff, Arizona, had compiled a list of some two-score nebulae that showed red shifts, and Hubble compiled a corresponding list of their distances. Developments in the general theory of relativity suggested that distant objects should in fact be moving away from us at a speed proportional to their distance, and in 1927 we found such a correlation—ragged, it is true, but showing a distinct increase of velocity with increasing distance. Hubble and Milton L. Humason at the Mount Wilson and Palomar Observatories extended this correlation to vastly greater distances; the most distant nebulae observed so far seem to be traveling at one fifth the speed of light and more.

Knowing the distances and velocities of the nebulae, it should be easy, by extrapolating backward in time, to tell when they were all together. And so it would, if we could be sure that these velocities have been the same from the beginning, or even

if we knew the law by which they have changed, if they have. But assuming for the moment that the velocities have kept the same, and adopting the new distance scale proposed last year by Walter Baade of Mount Wilson and Palomar, the answer is that the "origin" of all this motion lies five and a half billion years in the past. Is this time enough for the chemical elements to have evolved, in the proportions we now observe, from elementary particles? Is it long enough for stars and nebulae to have been built from primordial matter? For the creation of the solar system in all its parts? Such are the questions that cosmology asks of its sister science cosmogony.

But are we justified in assuming that the relative velocities have not changed during all these celestial eons? Probably not, for we should expect the various nebulae to interact, and such interaction could very well lead to changes in relative velocity. Specifically, the retarding pull of gravitation should slow the nebulae down. If gravitation is the only force effective between nebulae, then we should expect that the time scale inferred from the present slowed rate of expansion must be shortened—but by how much? At this point we have need of a theory, of some hypothesis on the dynamics of the universe, to point to further observational tests.

A theory opportunely presented itself, in the form of Einstein's general relativistic theory of gravitation. Hardly had this revolutionary extension of Newton's law been formulated when it was applied, by Einstein himself and by the Dutch astronomer Willem de Sitter, to the cosmological problem. Within a decade research leads converged from various directions upon an idealized model, or rather a series of models. The general theory of relativity sets out with few a priori prejudices concerning the geometry of space, holding the geometry to be in large measure contingent on the material content. This results in the intriguing possibility that physical space may be curved into a non-Euclidean form—a closed but unbounded universe of finite volume if the curvature

is positive, an infinite open one if the curvature is zero or negative. Observation should be able, in principle, to determine which obtains, and how great is the curvature. Thus if matter is distributed at random, the total number of nebulae seen out to a given astronomical distance should increase faster with distance in a closed space than in the flat Euclidean model, and the discrepancy should lead to an estimate of the amount of curvature.

Besides the models arising from Einstein's theory of gravitation, other relativistic models are possible, and several have been advanced as better suited to the facts. Among them is one in which the universe is stationary in time, but which yet allows for the general expansion. This it can do only by breaking with the postulate of conservation of matter and replacing it with the hypothesis of continuous creation of matter—the theory espoused by Fred Hoyle and his University of Cambridge colleagues.

As of the present no one model can lay exclusive claim to being the best representation of the actual universe. The choice sways from one to another as we choose to emphasize now one, now another set of partial observations, or as new horizons bring new knowledge. But the faith of science in the rule of law and the uniformity of nature bids us continue the search, confident that if we ask the right questions, and as we produce the means to answer them, all the parts of the puzzle will fall together into a consistent picture of the universe which portrays truth in the only sense in which science can sanction the word.

So history has shown. The world of ancient man was closed within a cramped sphere to which the stars were attached. The giants of the age of Newton pushed back this puny sphere to make room for the sun and its entourage, and their followers made of the stars a galaxy of majestic size and structure. And now in our day man leaps from it into the dim reaches of a universe of galaxies, groping with the problem of its structure, of whence it came, and of whither it goes. Truly one can say with Hubble: "The history of astronomy is a history of receding horizons!"

PART 2 THE SUBSTANCE OF THE UNIVERSE

THE ORIGIN OF THE ELEMENTS

by William A. Fowler

As a member of the W. K. Kellogg Radiation Laboratory at Cal Tech, the author, who is now professor of physics, got interested in the origin of the elements because of his wartime association with an astronomer. "During World War II," Fowler explains, "we worked on rocket ordnance, and one of the faculty members associated with us was the astronomer I. S. Bowen, who directed all photographic measurements in our field-testing program. After the war Dr. Bowen became director of the Mount Wilson and Palomar Observatories, and early in 1946 he held a series of informal seminars in his home on nuclear problems in astrophysics and astronomy. C. C. Lauritsen, director of the Kellogg Laboratory, and I and our students attended, and as a result of the interest that developed we decided to study experimentally in the laboratory those particular nuclear reactions which were thought to take place in stars. Research along these lines has been a part of our laboratory program since that time. In 1948 Jesse L. Greenstein came to Cal Tech, and his interest in the abundance of the elements in stars has stimulated much of our work."

THE ORIGIN OF THE ELEMENTS

by William A. Fowler

In investigating the nature and history of the universe we can hardly do better than to begin by examining what it is made of. The universe we see and measure is composed of an orderly yet diverse system of elements, from hydrogen to uranium. How did these elements come into being; from what primordial stuff were they made? As rare Ben Jonson shrewdly observed more than 300 years ago in *The Alchemist* (in a quotation to which the physicists Ralph A. Alpher and Robert C. Herman have previously called attention):

> *Ay, for 'twere absurd*
> *To think that nature in the earth bred gold*
> *Perfect i' the instant: something went before.*
> *There must be remote matter.*

Research into "remote matter" and the origin of the elements is going forward today along many paths, and of these none has been more fruitful than the study of the relative abundance of the various elements in the universe. The present abundances of the elements offer one of our most powerful clues to the history of the earth, the stars and the galaxies, for the abundance curve is the product of that history and was shaped by cosmic events. From this curve we can learn much about the evolution of stars, about cosmology and about all the grand-scale subjects of modern science.

Our inquiries into the composition of the universe are severely handicapped, to be sure, by the fact that gravitation, which acts alike on heavenly bodies, apples and human beings, has so far

chained mankind to his native planet. But notwithstanding this handicap, an imposing range of information on the universal abundance of elements is available to us today. There is, first of all, our own planet, where we can analyze at first hand the composition of the crust, oceans and atmosphere, and, allowing for losses of matter to space and redistribution of matter to the interior, can compute the proportions of the elements in the earth when it was formed. Second, there are the meteorites plucked by the earth from outer space; we attach considerable weight to these samples, because the matter in meteorites is assumed to have undergone less change than that in the earth's crust. Third, the light from a star, when analyzed with the spectroscope, identifies the elements on its visible surface. Every element emits or absorbs a characteristic spectrum of light (bright or dark lines at certain wave lengths) when its atoms are excited to high temperature; the elements have been "fingerprinted" in this way in laboratories, and their prints can be matched to the spectral light from stars. The abundance of each element can be estimated from the intensity of its radiation or from the amount of radiation the surface atoms absorb from the star's background radiation. Fourth, from galaxies and from interstellar space we can hear a song of hydrogen, in the form of radio waves at the 21-centimeter wave length; as radio astronomy develops it may tell us much more about the abundance of the elements in space. Finally, the cosmic ray particles that continually bombard the earth also supply us with samples of matter from the universe outside our planet.

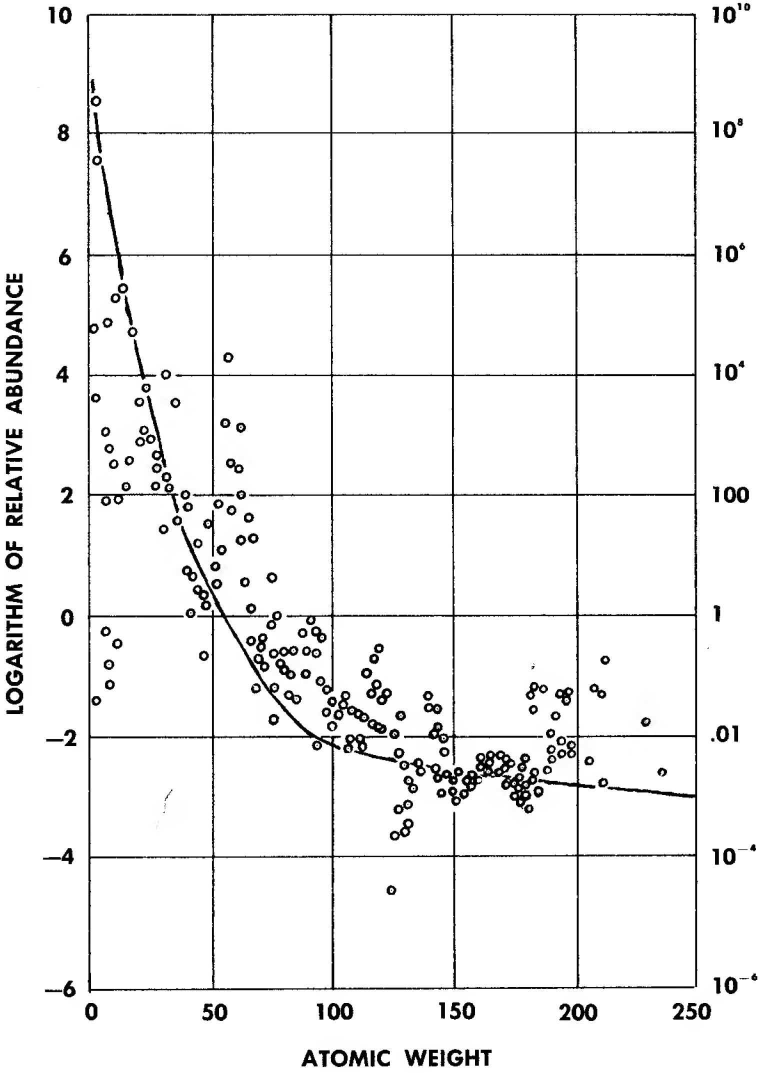
All these clues are beset with complications that may mislead us. Nor can we be confident that we have a true sample of the whole universe, for the information comes mainly from our own galaxy—indeed, largely from our own solar system. But it has been gratifying to find that every one of our methods of observation, when carefully carried out and corrected from complicating factors, yields much the same story. They produce a reasonable

and consistent picture of the average abundance of the elements in the universe as far as we can observe it. This picture—a curve showing the proportions of the various elements in the cosmos as a whole—is well represented by the curve constructed by Harrison Brown of the California Institute of Technology on the basis of his analysis of meteorites and other evidence (see chart on page 20).

By far the most abundant element is hydrogen: it accounts for 93 per cent of the total number of atoms and 76 per cent of the weight of the universe's matter. Helium is next: about 7 per cent by number of atoms and 23 per cent by weight. In general the abundance of the elements drops off with increasing atomic weight. The fall in the curve has one sharp interruption when we come to the elements of the iron group: these are about 10,000 times more abundant than their neighbors in the atomic-weight sequence. But except for this anomaly there is a general decline, and the heaviest elements add up to only a hundred millionth of all matter by number of atoms and a millionth by weight. It is a striking fact that all the elements beyond helium together amount to only a little more than one per cent of the mass of the universe.

If we take this picture to be correct, we have, then, a universal pudding composed of certain known ingredients mixed in certain proportions. Our task is to determine what recipe could have brewed this mixture.

We begin with the fact that, to the best of our knowledge, all the elements are made up of just two nuclear building blocks—protons and neutrons. (How the protons and neutrons themselves were created is a question outside the province of this article: only men of strong convictions, religious or scientific, have the courage to deal with the problem of the creation.) In a sense protons and neutrons are merely different versions of a nucleon: a free neutron may decay into a proton by shedding a negative electron, and the positively charged proton may become a neutron by combining with an electron or by emitting a positron.



Relative cosmic abundance of the elements is plotted. The solid curve itself is based on theoretical abundance expected from neutron-capture cross sections (see page 23). Anomalously high values near atomic weight 56 are the iron group; anomalously low values near 10, rare lithium, beryllium and boron.

The nucleus of the simplest element, hydrogen, is a single proton. Nearly a century and a half ago the Englishman William Prout suggested that all the elements consisted of combinations of hydrogen atoms. We have learned that the situation is vastly more complicated, but essentially most of the modern theories make a similar approach. It is natural to start with the working hypothesis that the elements were built up from protons or neutrons or both as the units.

The difficulty lies in trying to picture how this build-up took place and how it could have proceeded through the whole sequence to produce all the elements in the periodic table. The positively charged protons repel each other, and it takes a large amount of energy to overcome this repulsion and force them close enough together to combine. Some combinations are highly unstable or nonexistent. Other combinations in the sequence are so stable and so strongly bound that it is difficult to see how they can be transmuted or built up to larger atoms by natural processes.

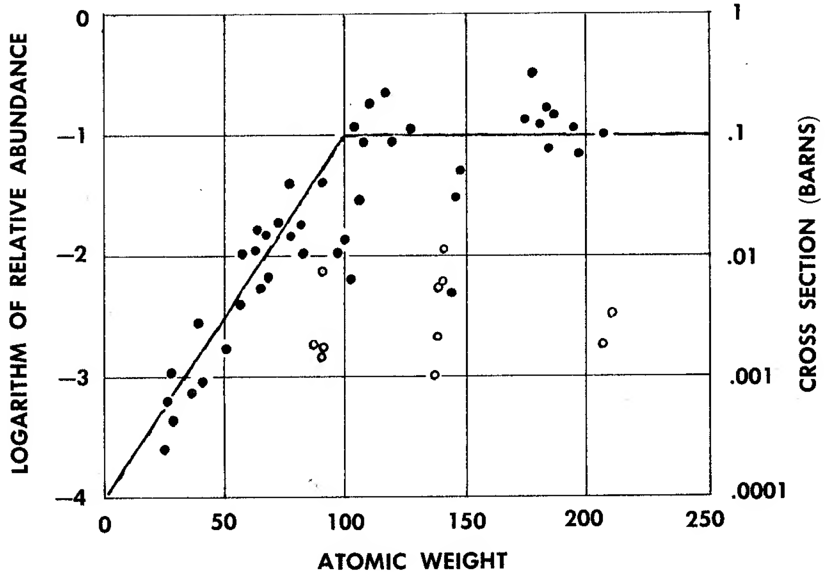
There are several current theories about the origin of the elements, but we shall consider only the two that have been worked out in fairly comprehensive fashion and are taken most seriously.

The more popular of the two is the one advanced by George Gamow and his collaborators. This theory holds that the elements were formed by a step-by-step build-up from neutrons. Gamow starts from the postulate, based on the apparent expansion of the universe, that the cosmos started from a core which exploded in a primordial "big bang" some five billion years ago. This exceedingly dense core, he believes, was made up primarily of neutrons, for under the great pressure electrons would be compressed into the protons. As the great neutron ball began to expand, some of the neutrons decayed to protons. Each proton promptly captured a neutron, the pair forming a deuteron, the nucleus of the hydrogen isotope of mass 2. Some deuterons then captured another neutron and became nuclei of tritium, or hydro-

gen 3. This nucleus soon decays by emitting a negative electron and thus is transmuted to helium 3. And so, by a rapid succession of neutron captures and electron decays, all the elements were built in the first burst of the universe's expansion. Gamow believes that the whole process of formation of the elements as we know them took place in a matter of a few minutes. The fleeing matter thereafter formed stars, planets and galaxies.

Two lines of evidence from laboratory experiments with particles give impressive support to Gamow's theory. First, it is well established that nearly all nuclei do in fact capture neutrons readily. Secondly, the neutron-capture cross sections of the various nuclei predict a pattern of element abundances which agrees remarkably well with the one actually observed. We should expect a simple relation between the neutron cross section of a given nucleus (i.e., the rate at which it captures neutrons) and the relative abundance of its production. Nuclei that capture neutrons rapidly should be comparatively rare when the sequence of element formation is completed, because most of them are quickly converted by such capture to other nuclei; conversely, nuclei that are slow to capture neutrons should accumulate to relatively high abundance. The curve of element abundances does in fact closely follow the curve of neutron-capture cross sections, in an inverse sense: that is to say, just as the curve of abundance falls sharply from hydrogen to the nucleus of atomic weight 100 and then flattens out, so the curve of neutron cross sections rises sharply from hydrogen to 100 and similarly flattens out beyond this atomic weight (see diagram opposite). There are even some correlations between fluctuations of elements from the two curves, notably at the neutron numbers 50, 82, and 126.

But there are important difficulties with Gamow's theory—difficulties to which his collaborators Ralph A. Alpher and Robert C. Herman have themselves called attention. The most serious is the fact that in the sequence of atomic weights numbers 5 and 8 are vacant. That is, there is no stable atom of mass 5 or of mass 8. We can produce helium 5 in the laboratory by bombarding



Cross sections of nuclei for neutron capture, as measured by Donald J. Hughes and his collaborators at the Brookhaven National Laboratory, are plotted here. The heavier nuclei capture neutrons more readily. The exceptionally stable nuclei are indicated by open circles. A barn (*right-hand vertical scale*) is 10^{-24} square centimeters.

helium 4 with neutrons, but it immediately breaks down to helium 4 again. Likewise we can produce momentarily an isotope of beryllium of mass 8, but it too instantly breaks down (by fission into two helium 4 atoms). The question then is: How can the build-up of elements by neutron capture get by these gaps? The process could not go beyond helium 4, and even if it spanned this gap it would be stopped again at mass 8. In short, if neutron capture were the only process by which elements could be built, starting with hydrogen, the build-up would get no farther than helium.

This basic objection to Gamow's theory is a great disappointment, in view of the promise and philosophical attractiveness of the idea. The other major current hypothesis is less simple and less elegant; it complicates the picture by invoking other processes, in addition to neutron capture, to account for the build-up of the elements. But it seems to surmount the difficulties encountered by the Gamow hypothesis.

The theory argues that the elements were built not in a primordial explosion but in the hot interiors of stars. It starts from our knowledge that nuclear reactions and transformations must be going on constantly in the stars. As Sir Arthur Eddington presciently remarked in 1920, after Lord Rutherford had transmuted nuclei by bombardment in his laboratory: "What is possible in the Cavendish Laboratory may not be too difficult in the sun." Eddington's informed guess was certainly correct, but not until 1938 was it translated into terms of specific processes. Hans A. Bethe, seeking to account for the enormous and enduring energy of the sun and other stars, conceived two chains of nuclear reactions that would explain their tremendous release of energy and would build new nuclei. The processes have been known ever since as proton-proton fusion and the carbon-nitrogen cycle. The new theory of synthesis of the elements, which has been championed most extensively by Fred Hoyle of the University of Cambridge, assigns key roles to these processes.

We start with a universe consisting of a cold, dilute and turbulent gas of hydrogen atoms. By gravitational attraction part of the gas condenses into stars. As a star contracts under gravitational force, its interior grows very dense and hot. When the central temperature reaches about five million degrees, the protons are moving with enough energy to fuse on colliding and form deuterons. Deuterons in turn combine with protons to form helium 3. Helium 3 does not interact with protons, but laboratory experiments have shown that two helium 3 nuclei can fuse and produce helium 4, ejecting the two surplus protons. The net

result of this proton-proton chain is the conversion of four atoms of hydrogen into one atom of helium.

In this way a core of helium develops in the center of the star and gradually grows in size. After a time, as the hydrogen fuel in the interior is used up, the core begins to cool. It then contracts, because gravitational forces gain the upper hand. As a result the temperature of the core rises again. The sudden rise of the internal temperature heats up the star's outer envelope of hydrogen; the mantel expands enormously; its extended surface then radiates cooler (i.e., redder) light, and the star becomes a "red giant."

We have now a star with a hot core of helium, at a computed temperature of more than 100 million degrees. What happens next? We have come to the Gordian knot of the speculations on the build-up of the elements. Two helium nuclei may combine to form a nucleus of mass 8, but as we have seen, any nucleus of mass 8 must be extremely unstable, for none is found in nature. However, beryllium 8 *has* been produced momentarily in the laboratory, and will certainly materialize in the very hot and dense interior of a star. In fact, in that environment beryllium 8 will be produced at as fast a rate as it breaks down, so that a small amount of it is always present. If so, an occasional beryllium 8 nucleus may during its very brief lifetime fuse with a helium 4 nucleus. The combination should result in a nucleus of carbon 12.

Hoyle has pointed out that, in view of the extreme rarity of the beryllium 8 nuclei (about one part in 10 billion in a 100-million-degree star), the beryllium 8 nucleus had better have a big cross section for capturing helium nuclei if this scheme is to work. Naturally the question cannot be put to a direct test by bombarding a beryllium 8 target in the laboratory, for the nucleus is too ephemeral. But in the W. K. Kellogg Radiation Laboratory at Cal Tech we have been able to obtain indirect evidence that this capture does have a high probability, or, in the parlance of nu-

clear physics, that it is a "resonant" reaction. Hoyle reasoned that if the reaction is indeed a resonant one, the product, carbon 12, must go through an excited state with certain specified properties. We have found that the carbon 12 nucleus can in fact take this excited form, with almost exactly the properties Hoyle predicted. We produced excited carbon by bombarding boron with high-energy deuterons. The excited carbon 12 nucleus resulting from this reaction promptly disintegrated into three helium nuclei. On the basis of very general physical principles we can argue that in the hot core of helium in a star the reverse process can take place: that is, three helium nuclei may combine to form excited carbon 12, which may then discharge its energy of excitation and become stable carbon.

The jump from helium to carbon of course skips the elements lithium, beryllium (whose stable form is beryllium 9) and boron. There is good reason to suppose that these elements are not produced in the main line of build-up of the elements. They are comparatively rare, and may be made by secondary processes. It is known, for instance, that bombardment of heavy elements with hydrogen nuclei sometimes chips off fragments which are identifiable as nuclei of lithium, beryllium and boron. Possibly this process goes on in spots ("sunspots") on the surfaces of stars or occurs in stellar explosions.

Once carbon 12 has been synthesized in the helium core of a star, it may build up by successive captures of helium nuclei to oxygen 16, neon 12 and perhaps magnesium 24. When the helium has been largely used up, so that there can no longer be much release of energy from these fusion reactions, the core cools and contracts. The contraction again raises the temperature of the core, this time perhaps to an energy high enough to trigger interactions among the nuclei of carbon, oxygen and neon. Such reactions would produce the silicon group of elements (around atomic weight 28). The temperature of the core may continue to rise until, at about five billion degrees, the build-up of elements by fusion reaches a dead end. At this stage the build-up would

form the most stable of all the elements, namely iron and its neighbors (around atomic weight 56). Any nuclear reaction involving the iron group must absorb energy rather than release it; hence these nuclei cannot serve as fuel to continue the chain of fusions.

Hoyle has suggested that this impasse may account for the anomalous abundance of the iron group of elements in the universe. As the primeval stars grow older, they accumulate iron as the end product. If they reach the stage where they have burned up all their internal fuel and then explode (perhaps as a result of a sudden disturbance of the hot core material and its reaction with unburned material in the envelope of the star), they will fling a considerable amount of iron into interstellar space.

We must now pause to relate the element-building processes to the evolution of stars. Clearly in the early stages of a star's evolution the only, or at least dominant, process is the build-up of hydrogen to helium. The fusion of hydrogen to helium is, in fact, the main source of energy of most stars (which fall in what is called the "main sequence" on the familiar chart of star classifications). Recall that most of the matter in the universe is hydrogen and helium: we can assign the building of all the other elements to comparatively minor or rare processes in the life of stars.

It is in the old "red giants" that the fusion of helium into carbon and successively heavier elements takes over the dominant role. But, as we have just seen, we have reached an impasse at iron, and we must now find some way to construct the elements beyond the iron group. Here Gamow's concept of build-up by neutron capture, and what we know of certain cataclysmic events in the history of stars, comes to our aid.

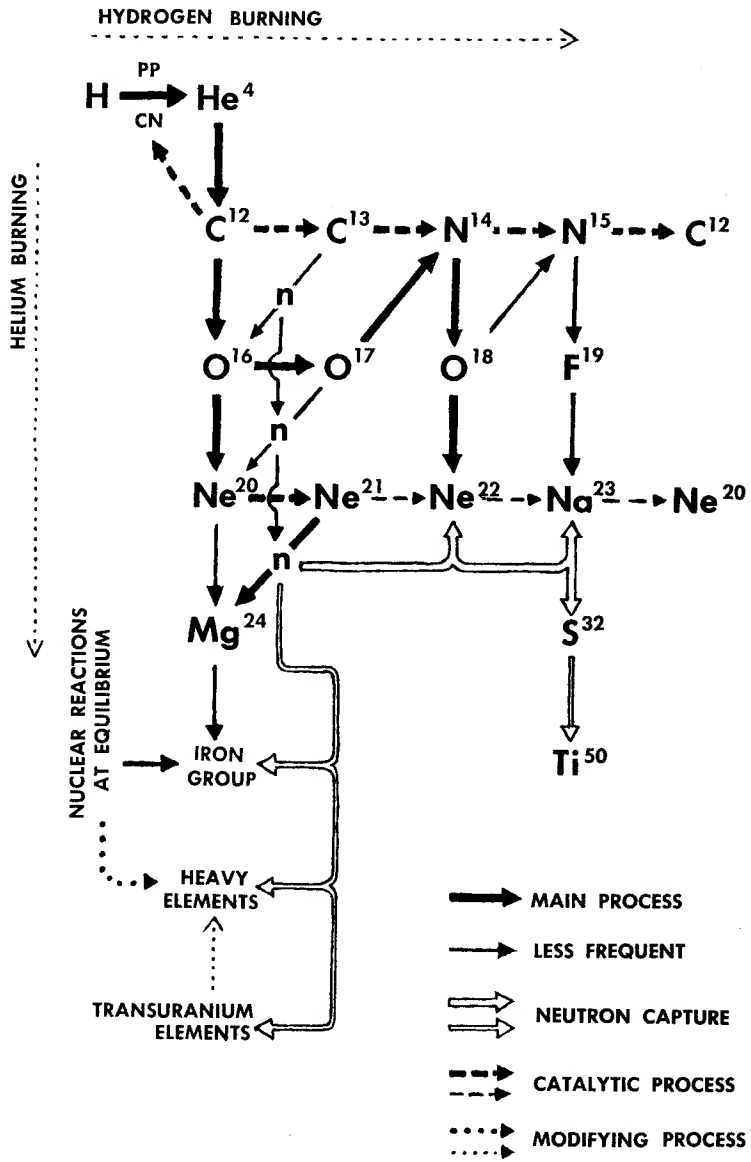
Stars, like human beings, are subject to accidents and disorders: not all of them live to a ripe old age. They occasionally boil up to a state of instability that results in their exploding as novae or supernovae. This may happen to a star of any age, young

or old. When a young star explodes, it discharges hydrogen and helium into interstellar space. An old star will spew forth not only these nuclei but also other elements from carbon up to iron. Besides this, even stable stars, including our sun, are known to be constantly ejecting corpuscles of matter into space.

Thus a debris of matter from living and dying stars pours into space, and its elements mix with the interstellar gas. From this material new stars are born: astronomy today has strong evidence of the existence of young or infant stars in the heavens. So we can postulate two kinds of stars: primeval or "first generation" stars, and "second generation" stars, which start with a legacy of the elements up to iron from the parents of their matter.

Let us now consider a second-generation star which has condensed from hydrogen mixed with some carbon, oxygen, neon and even a little iron. In these stars hydrogen in the core will again be converted to helium, but now, because carbon is present, the conversion will take the route of the second process described by Bethe, the carbon-nitrogen cycle. In this cycle carbon 12 captures hydrogen nuclei in a series of steps which converts it successively to carbon 13, nitrogen 14 and nitrogen 15: in the end nitrogen 15 takes on another proton, breaks down to carbon 12 again, and in so doing emits a nucleus of helium. Thus the chain of reactions

Entire scheme of the synthesis of elements in stars is presented on the opposite page. Elements synthesized by interactions with protons (*hydrogen burning*) are listed horizontally. Elements synthesized by interactions with alpha particles (*helium burning*) and more complicated processes are listed vertically. The letters *pp* stand for the proton-proton reaction; the letters *CN*, for the carbon-nitrogen cycle. The letter *n* stands for neutrons liberated in nuclear reactions and thus available for neutron-capture processes. The production of carbon (C), nitrogen (N), oxygen (O), fluorine (F), neon (Ne), and sodium (Na) are given in detail. Described in less detail are the neutron-capture processes responsible for magnesium (Mg), sulfur (S), titanium (Ti), the iron group, the heavy elements and the transuranium elements. The broken line between the transuranium elements and the heavy elements represents alpha decay or fission.



produces helium and all the isotopes of carbon and nitrogen. It can be calculated that this process, rather than the direct fusion of protons, is the source of energy in second-generation main-sequence stars which are large enough to have internal temperatures over 15 million degrees.

The oxygen in the star's core mixture is converted by proton capture to the isotope oxygen 17, and neon similarly to neon 21. Now these isotopes, and carbon 13, come to play a crucial role when the star arrives at the red giant stage and its core consists mainly of hot helium. The three isotopes, on reacting with helium, produce unstable nuclei which emit neutrons; so laboratory experiments have shown. Consequently they furnish a steady supply of neutrons within the core. We have seen that all nuclei, even iron, readily capture neutrons. Here, then, is the mechanism that breaks the iron bottleneck. By successive captures of neutrons, nuclei can be built up from the iron group to elements as heavy as lead and bismuth. The slow neutron-capture process in the core of a star cannot carry the build-up beyond bismuth, because the heavier elements decay too rapidly (by emitting alpha particles, or helium nuclei). However, the heavy elements *can* capture neutrons at a sufficiently rapid rate to continue the chain during an explosion of a star.

That the stars do in fact synthesize heavy elements has been confirmed by considerable evidence, some of it spectacular. The most dramatic was the discovery of the element technetium in certain giant stars (through its spectral fingerprints). Technetium is an unstable element whose longest-lived known isotope has a half life of only 216,000 years—far less than the age of the stars in which it is found. It must therefore have been made in the star long after the star's birth. As for the synthesis of the heaviest elements, an isotope of the element californium was found in the debris from a thermonuclear explosion in the Bikini tests of 1952, and we have seen an intriguing suggestion of its presence in certain supernovae. After their original flare-up, these exploding stars decline in brightness at a rate which corresponds

to a half life (decline to half the intensity) of 55 days, and this is just the half life of spontaneous fission of californium 254!

Research in our own and other laboratories has now established possible processes for synthesis of all the elements (see page 29). Of course this scheme is still highly tentative. It is disconcerting that so many different processes have to be invoked; it would be much more satisfying to see a single process that could build all the elements. The picture may, however, become simpler as more research is done. What particularly gratifies workers in this field is that speculation about the origin of the elements has been reduced to questions specific enough to be tested both by nuclear physicists in the laboratory and by astrophysicists studying the stars.

There is food for philosophical thought in what has been learned so far. The heavy elements, of which our solar system has its full share, took a long time to produce—probably one to two billion years. Thus the particular part of the universe we inhabit is not the oldest thing in it; many cosmic events preceded the formation of the earth. The oldest stars in our galaxy are estimated to have an age of 6.5 billion years, while analyses of meteorites indicate that the solar system is no more than 4.5 billion years old.

Copernicus displaced the center of the universe from the earth to the sun; later cosmologists dethroned the solar system as the center; now we see that our system was not even in existence at the beginning of the galaxy. So dies the last vestige of mankind's geocentric conception of the universe.

PART 3 THE FORM OF THE UNIVERSE

I. THE CONTENT OF GALAXIES
by Walter Baade

Walter Baade is astronomer at the Mount Wilson and Palomar Observatories. He was born in Schroetinghausen, Germany, in 1893 and attended the universities of Münster and Göttingen, receiving his Ph.D. in 1919. For 12 years he was on the staff of the Hamburg Observatory and towards the end of that period was also *Privatdocent* at the University of Hamburg. He came to Mount Wilson in 1931.

II. THE EVOLUTION OF GALAXIES
by Jan H. Oort

The author is professor and director of astronomy at the University of Leiden in the Netherlands. He entered the University of Groningen in 1917 to study physics or astronomy, "two subjects by which I had been fascinated during my high-school years. There I soon came under the inspiring teaching of J. C. Kapteyn, who was one of the great pioneers of galactic research. This determined my further scientific life. From the beginning I was particularly attracted by problems of stellar dynamics, both for our own galactic system and for other systems, but always only insofar as the problems were directly connected with observations. This included, of course, the study of the interstellar medium, and now also includes radio astronomy. Besides, I worked on the Crab Nebula and on the origin of comets. I am intensely interested in the unknown that lies before us in radio astronomy. It will certainly teach us a great deal more about the present state of the universe. I am hoping that it may give us some observational insight into the beginning of the universe. We may also hope to begin to understand the origin of the spiral structure of galaxies. But it is so often the unexpected which is the most important."

III. COLLIDING GALAXIES
by Rudolph Minkowski

One of the first astronomers to find a visible radio "star," Rudolph Minkowski is a member of the staff of the Mount Wilson and Palomar Observatories and a research associate at the California Institute of Technology. He was born in Strasbourg and studied physics at the University of Breslau. After receiving his doctorate, he joined the faculty of the University of Hamburg in 1922. There his scientific interest, always rather close to astronomical problems, moved more and more into astronomy, and he concentrated on investigating the intensities and widths of spectral lines, particularly of the light emitted by nebulae. He went from Germany to Pasadena in 1935. There his work has been devoted mainly to supernovae, planetary nebulae and more recently to radio sources in the sky.

THE CONTENT OF GALAXIES
by Walter Baade

During the last 30 years, ever since the discovery of the apparent expansion of the universe, studies of our cosmos have been dominated by the cosmological approach: the attempt to understand the structure of the universe in terms of the geometry of space and to estimate its age from its expansion. But as William A. Fowler has made clear in the preceding chapter, we can also explore these questions by studying the matter the universe contains. After all, the long history of the universe must have left its marks on the material we now observe, and we should be able to obtain information about this history—specifically the formation of stars and galaxies—by examining the composition of these systems. Within the past 10 years we have come to realize that there is an intimate relation between the structure of a galaxy and the character of its population of stars, a relationship which gives a clue to the origin of stars and to the evolution of galaxies.

In order to appreciate the significance of this relationship we must first have a brief look at the forms of galaxies and their classification. This classification was established by the late Edwin P. Hubble of the Mount Wilson and Palomar Observatories. Hubble showed that the galaxies can be divided into three broad groups. The first comprises systems which are spherical or nearly spherical in shape and are filled with stars, apparently packed densely at the center of the galaxy and gradually thinning out to the edge of the system. These galaxies range in shape from truly spherical to ellipsoidal forms. Because their projection on a photographic plate usually has an elliptical outline, Hubble called this group the "elliptical" galaxies. About 17 per cent of all the brighter galaxies observed fall in this class.

The second class, comprising the majority of the galaxies studied (about 80 per cent), have a spheroidal body at the center, but in them a flat disk surrounds the central body. Spectroscopic observations show clearly that the disk rotates rapidly around the axis of the system. Because the disk usually contains spiral arms, this group are called spiral galaxies. Hubble distinguished three subclasses among the spiral galaxies: the so-called Sa spirals, in which the spheroidal body is so large that it almost envelops the outer disk; the Sb spirals, in which the central system is considerably smaller than the disk; and the Sc spirals, in which the central body has shrunk to a bright kernel of almost insignificant size. A peculiar type among the spiral galaxies are the so-called "barred" spirals, in which the disk has degenerated into a broad bar running through the center, the spiral arms emanating from the two ends of the bar.

The third class are the irregular galaxies, which show no recognizable form or order, except that most of them seem to be flattened. The Clouds of Magellan are typical examples of this class. Only 2 to 3 per cent of the galaxies belong to the irregular group.

Hubble made an extensive photographic survey of all the galaxies in the Northern Hemisphere above a certain brightness, and two important internal differences between spiral and elliptical galaxies emerged. One was the fact that the spiral galaxies as a rule contain large quantities of dust, whereas the elliptical systems rarely do. We shall consider the significance of this observation later. The second observation, which became the starting point for a long and fruitful investigation, was that while the spiral galaxies were easily identifiable as collections of stars, not a single elliptical galaxy could be resolved into individual stars!

Intensive studies were made of the galaxies closest to our own. The Milky Way system is a member of a cluster of 17 galaxies

concentrated within a radius of about one million light-years; outside this cluster we have to travel eight million years before we encounter the next spiral galaxy. Nearly all types of galaxies are represented in our local group. For illustration let us consider a set of examples of different types which are all at nearly the same distance from us. An irregular galaxy (IC 1613) and an open, Sc spiral (NGC 598) proved to be easily resolvable into stars throughout their extent; in fact, even a 20-inch telescope resolved the brightest stars in these galaxies. The case of the Great Nebula in Andromeda, a spiral of the Sb type (with a large central spheroid), was quite different. Only the spiral arms of this galaxy were resolvable: in the central body and the regions between the spiral arms no individual objects could be distinguished. Finally, the galaxies of the elliptical type, the two companions of the Andromeda Nebula, were not resolvable at all, in spite of repeated attempts to force their resolution.

What did these findings mean? Evidently, if the unresolved systems were made up of stars, these stars were too faint to be distinguishable individually by our most powerful telescope at that time—the 100-inch on Mount Wilson. This meant that any stars in the unresolved parts of the Andromeda Nebula, for example, must be at least 100 times fainter than the brightest stars in its spiral arms. Now there were a number of good reasons to believe that the central part of the Andromeda Nebula was made up of stars. One of them was the fact that novae, known to signal the explosion of stars, occur in the central region of that galaxy. What kind of stars might be hidden there? The only clue at the time was that they are at least 100 times fainter than the brightest stars in the spiral arms of the Andromeda Nebula. It was possible to determine the intrinsic, or absolute, brightness of those stars, because the Nebula's distance from us was known. It turned out that the Andromeda Nebula's brightest stars are of the same luminosity as the brightest stars in the surroundings of our sun. Like the brightest stars in the neighborhood of the sun, they are

very blue stars. In short, the stellar population of the spiral arms of the Andromeda Nebula is very similar to the population of stars around us.

All this was well known by the early 1940s. It left the stellar populations of the elliptical galaxies and of the central bodies of the spiral galaxies and of the central bodies of the spiral galaxies as mysterious as ever. Only this much was clear: the stars were fainter than the faintest objects the 100-inch telescope was then capable of recording. Obviously the problem had to wait until the 200-inch telescope came into operation.

When the outbreak of the Second World War postponed the completion of the 200-inch telescope indefinitely, I decided to try once more at the 100-inch. I knew it would call for every trick of the game. But wartime conditions brought certain advantages: I could get all the time at the telescope I wanted and could thus wait for good observing conditions, and the blackout of the Los Angeles valley restored the dark night sky of the early Mount Wilson days and made it possible to utilize the high sensitivity of modern photographic plates.

The new attempts to resolve the central region of the Andromeda Nebula began in the fall of 1943 on blue-sensitive plates of high speed. In the very last plate of this trial run we found unmistakable signs of a beginning of resolution. No stars had yet emerged, but the formerly smooth appearance of the unresolved area had broken up into a curious patchiness. (As we learned later, this reflected small-scale fluctuations of the underlying star field.)

No further gains could be expected from blue plates, because they had already been exploited to the limit of their sensitivity. The only alternative was to use red-sensitive plates, whose speed had been improved, and long exposures. No marked gain could be expected from the red plates unless the brightest stars in the unresolved systems were red, but luckily one could predict that they must be red indeed. Photoelectric measures had shown that

the elliptical galaxies and the central systems of spirals are distinctly reddish in color.

The exposure time necessary to force the resolution of our systems turned out to be four hours. At this point a new difficulty arose. In four hours the focus of the 100-inch telescope may change by a millimeter or more because of the cooling of the secondary mirror, which is exposed to the cold night air. But it was absolutely essential to stay within one tenth of a millimeter of the actual focus, because the star images were certain to be extremely close together. Therefore a method had to be devised to adjust to the focus changes with the prescribed accuracy. Such a method was perfected, and in the fall of 1944 everything was ready for the final attack.

The new campaign was a complete success. It led in rapid succession to the resolution of the Andromeda Nebula, of its two companions (the elliptical systems M 32 and NGC 205) and of the elliptical galaxies NGC 147 and NGC 185. There could not be the slightest doubt that real resolution into stars had been achieved. Comparison of different plates of the same system showed that for each star present on one plate there was a corresponding stellar image on the other. And indeed, in the elliptical galaxy NGC 205 we actually identified some variable stars (later confirmed with the 200-inch).

What kind of star population emerged on the plates? The brightest stars in the newly resolved systems turned out to be red giants of the same kind as the brightest stars in globular clusters in our galaxy. Exhaustive analysis of several kinds of evidence made this certain. Furthermore, the elliptical galaxies were found to share another distinctive feature of the globular clusters: they contain many short-period Cepheid variable stars.

Thus it was clear that elliptical galaxies and the central bodies of spiral systems have the same general stellar make-up as globular clusters. Their stars are quite different from those in the neighborhood of our sun and in spiral arms. The latter form a sequence whose brightest stars are blue supergiants; the stars of

the globular clusters and elliptical systems are of a class whose brightest members are red giants. The stars making up the contrasting systems could therefore be distinguished into two different populations. They were named Population I (as in spiral arms) and Population II (as in globular clusters).

Obviously all elliptical galaxies are pure Population II systems. In spirals with central spheroidal systems, both populations are present—Population I in the spiral arms, Population II in the central body (see Plates 12 and 13). In spirals without a large central core and in irregular systems the splash of Population I is so overwhelming that the much less conspicuous Population II is clearly obliterated. For a while it was believed, indeed, that the irregular Clouds of Magellan were pure type I systems, but short-period Cepheid variable stars and novae (both Population II types) were found in them. There is strong evidence that all galaxies have at least some Population II (see Plates 15 and 16).

What is this rather inconspicuous but obviously ever-present Population II? The answer is: a population of very old stars, in fact the very oldest of which we have any knowledge so far. They are known to be about five billion years old.

When I examined the first red exposure of the Andromeda Nebula that resolved the central region, I was very surprised to discover two large clouds of luminous gas in a spiral arm which happened to cross the field. I had previously photographed the same region on blue-sensitive plates, and those plates had not shown the clouds. Moreover, Hubble, in his earlier survey with blue-sensitive plates, had been unable to find a single luminous nebulosity in the whole Andromeda Nebula! Obviously blue plates simply fail to detect these objects (called "emission nebulosities"), while red-sensitive plates do. To get a clearer picture of the situation I made a new survey of the Andromeda Nebula on both red and blue plates. The result was the discovery of nearly 700 emission nebulosities on the red-sensitive plates. They show a most striking arrangement—strung out like pearls along the spiral arms. This restriction to the arms is not surpris-

ing, because such a nebulosity must be excited by a hot star of the Population I type, and these stars occur, as we know, only in the spiral arms.

The survey confirmed that blue plates register these luminous clouds only faintly if at all. At the same time N. U. Mayall of the Lick Observatory showed by analysis of the clouds' spectra that they emit exactly the same kind of light as do luminous clouds in our own galaxy. There was only one explanation: namely, that the blue part of the light from these clouds in the Andromeda Nebula was heavily absorbed by dust in the Nebula. Dark lanes along its spiral arms confirm the high density of dust in the arms. On the other hand, there is good evidence that outside the arms the density of the dust is much lower. The Andromeda Nebula is surrounded by more than 200 globular clusters. If it were filled with dust throughout, light coming to us from clusters on the farther side of the Nebula would be heavily reddened. No reddening comparable to that in the emission nebulosities has been found in these clusters. The few heavily reddened globular clusters found in the Andromeda Nebula are located in the spiral arms.

Altogether there are good reasons to believe that the most basic feature of the structure of spiral galaxies is the dust and gas in their arms, and that the Population I stars represent a secondary phenomenon. To put it differently: it is the dust that makes the stars, and not vice versa. The nuclear processes responsible for the energy of stars offer a very cogent argument in favor of this conclusion. The hot stars of the spiral arms burn their fuel so fast that their lifetime hardly can exceed 50 to 100 million years. The continued presence of such stars in spirals wherever we look means that burned-out stars must be continuously replaced by new ones. During the last few years investigation of the so-called "stellar associations" of our own galaxy has made us acquainted with whole groups of young stars barely older than a few tens of millions of years. All are typical Population I stars. There cannot be the slightest

doubt that all Population I stars are young stars, cosmically speaking.

We are now able to outline in bold strokes at least a part of the history of the galaxies. Some five billion years ago there was a big burst of star formation in all galaxies. In the spheroidal galaxies, which rotate slowly, apparently all the available dust and gas were formed into stars at that time, for we find in them today only old (Population II) stars. (That their supply of dust and gas is largely exhausted is documented by Hubble's finding that cosmic dust is exceedingly rare in elliptical galaxies.) In all rapidly rotating systems, which had flattened into disks, only part of the dust and gas was converted into stars. The remaining part formed the spiral structure of the disk and has produced stars ever since.

THE EVOLUTION OF GALAXIES

by Jan H. Oort

SURVEYING the galaxies in the heavens, astronomers in the past couple of centuries have been in a position like that of a lookout watching the approach of an armada of strange objects. The objects came into view first as dim, fuzzy forms ("nebulae"). As more powerful telescopes brought them closer and closer, they were identified as collections of stars, then distinguished into systems of varied shapes and types; today we can resolve the details of internal structure in many of them. With this clearer view of other galaxies, man has acquired a new perspective on the galaxy in which he lives—the Milky Way—and can begin to speculate intelligently about the origin and evolution of galaxies in general.

Several thousand galaxies (of the billion or so visible with present telescopes) are close enough to show details of their structure. They vary widely in shape and texture. At one extreme are chaotic, mottled-looking systems which have been named "irregular" galaxies (Plate 11); at the other are perfectly smooth, symmetrical systems called "elliptical" galaxies (Plate 10). Between these extremes is a great variety of systems; most of them, including our own Milky Way, have a spiral structure with wide-flung arms encircling the central nucleus of the galaxy (Plates 2–7). The majority of galaxies examined are spirals. Elliptical systems make up roughly 20 per cent of the total and the irregular systems comprise some 2 or 3 per cent.

On its face, this general picture suggests that the evolution of galaxies may proceed from the irregular through the various spiral types to the smooth, elliptical form as the final stage. But when we look at the picture in detail it becomes difficult to see how the

spiral systems could evolve to the regular, more or less globular shape of the elliptical galaxies. This chapter will present a theory attempting to explain the evolution of spiral and elliptical galaxies.

The spiral galaxies are particularly intriguing, for a number of reasons. We shall consider their structure in detail, with the help of the photographs in Plates 2–7, which illustrate their varying features and forms. Galaxies are usually identified by number, prefaced by initials for the source from which they got their names. The brightest have the initial M (e.g., M 51) for Charles Messier, the French astronomer who listed the 103 brightest nebulae in 1784; many of the fainter ones are designated by NGC, for the New General Catalogue of the Danish astronomer John L. E. Dreyer, which was based largely on discoveries of nebulae by the eighteenth- and nineteenth-century English observers William and John Herschel—father and son—who pioneered the study of galaxies.

The spirals take many forms. Some have two principal arms, stemming from opposite sides of the core of the galaxy. Sometimes the galaxy has a large, smooth central mass, like an elliptical galaxy, with spiral arms wound tightly around it. About 30 per cent of the spiral systems are so-called "barred spirals"—formations of bars across the central region with spiral arms projecting from their ends in various circular or open forms.

Practically all the spiral galaxies possess an important feature in common: they have a flat, wheel-like shape and the arms lie in this plane. This flatness is clearly evident in spiral systems that we see edge on. The most striking example of all is our own galaxy: it is a disk some 80,000 light-years in diameter from edge to edge but only about one hundredth of that distance in thickness.

The spiral galaxies have another common characteristic. Like the irregular galaxies (but unlike the elliptical ones) they all have great clouds of gas. Their spiral arms show bright patches that can be identified as luminous gas; beside this we can detect many clouds which, though cool and invisible, betray their presence because they contain dust and block light. Dark bands and streaks defining such clouds appear in spiral galaxies seen edge on and along the arms of spirals seen face on.

During the past four years the clouds of gas in our own galaxy have been mapped from our station in the Netherlands by means of radio receivers tracing the 21-centimeter radiation from hydrogen gas in the clouds. Most of the clouds are confined to a thin, flat section along the central plane of the Milky Way; in fact, the faint luminescence of the Milky Way comes from distant star clouds in this thin layer. The gas is not spread evenly throughout the plane but rather is concentrated in long lanes. There cannot be much doubt that these lanes define spiral arms, such as we see in other galaxies. The spacing between the lanes (roughly 6,000 light-years) is a measure of how tightly the spiral arms are wound around the core of the galaxy.

These observations confirm what studies of other galaxies have led us to suspect: namely, that the interstellar gas in spiral systems is largely concentrated in the arms. The radio measurements also tell us a good deal more. They prove that the galaxy is rotating around its center, and that the arms are trailing. And from the Doppler effect on the wave length of the 21-centimeter radiation we can calculate the speed of rotation of different parts of our system. It turns out that the galaxy is not turning uniformly around the center. Our own solar system and its neighbors, for instance, take about 230 million years to make a complete revolution around the center of the galaxy, but the stars and clouds nearer the hub (at half the distance from us to the center) travel around it in only 120 million years.

The measurements of speeds of motion in the galaxy give us important information on the dynamics of the system, which in turn has a crucial bearing on the galaxy's origin and evolution. The velocity of rotation indicates the strength of the centrifugal forces acting on the spiral arms. Since the centrifugal force must be balanced by gravitational force in any system that holds together, the rotational velocity also gives us a means of estimating

the mass of the galaxy as a whole and of its various parts. The total mass of our galaxy turns out to be 70 billion times that of the sun. Of this total about 94 per cent is accounted for by the stars and only about 6 per cent by the interstellar gas.

The arms of the spiral galaxies have always been a major enigma. How can they keep their form and structure as the galaxy whirls rapidly in space? One would suppose that the shearing effect due to the differing rates of rotation of different parts of the galaxy, as well as internal motions within the arms themselves, would wipe out the arms before the galaxy had made many revolutions.

It seems unlikely that gravitational attraction alone could form an immense, tenuous arm of this kind or hold it together. To explain the existence of the spiral arms we must look to other forces, and, as we shall see, these forces involve the clouds of gas in the arms.

There is another striking feature of the structure of spiral galaxies that must be explained by any theory of their origin and history. As Walter Baade has explained in the preceding chapter, spiral galaxies are made up of two types or "populations" of stars: Population I consists of supergiant blue stars and gas clouds; Population II is distinguished by red giant stars and contains little gas. Now these populations have very different distributions in the galaxy. Population I is concentrated within a thin, flat disk extending over the plane of the galaxy. Stars and clusters of the Population II type, on the other hand, have an almost spherical distribution in the galaxy and are strongly concentrated toward the center. There are classes of Population II stars which, like Population I, are found only within a flat disk in the plane of the galaxy (we call them "disk Population II"), but these stars, unlike the irregular Population I, have a smooth distribution and are densely concentrated around the center.

How are we to explain this curious coexistence of different populations, distributed in very different patterns, within a

galaxy? This question, and the others we have already considered, can be answered if we postulate the following theory of the origin of the various types of galaxies.

Let us suppose that the universe consisted at first of a thin, expanding gas, denser in some places than in others. At a certain stage in the expansion of the universe the internal gravitational force in some of the denser regions apparently became sufficient to overcome the velocity of expansion with which all matter had been endowed from the beginning. These aggregations of matter detached themselves from the rest of the universe, began to develop as independent units ("protogalaxies") and thus set off on the road to becoming galaxies. The protogalaxy was an irregular mass with large-scale random internal motions (as well as smaller turbulences); but for our lump of universe to stay together, the gravitational pull of the whole lump on its separate parts had to be strong enough to overcome the original relative velocities. After a time the mass of gas started to contract under the influence of this gravitation. As the gas currents collided and intermingled, some of their energy of motion was converted to heat, which was radiated away. The slowing of the gas's motions made it possible for the system to contract further, and at the same time to even out its irregular shape.

If these processes of contraction and loss of energy through radiation had continued indefinitely, our protogalaxy might have ended up as a small and very dense mass of gas. But two factors set a limit to the contraction. First, some of the denser lumps of matter within the mass contracted more rapidly than the rest and condensed into stars. The birth of stars put an end to contraction, because the stars, being much smaller than gas currents or clouds, no longer had much chance of colliding and losing energy of motion. The second limit on contraction would develop from rotation of the system. A system formed in the way we have described would inevitably start rotating as a result of the intermingling of currents in the original gas—as an eddy forms when

currents of water meet. Now as the system contracted, its rotation would speed up through the conservation of angular momentum (just as a whirling skater turns faster when he pulls in his arms). The increase in rotational velocity would increase the centrifugal force. Eventually, when the centrifugal force came to equal the gravitational attraction, the system would cease its contraction in the plane of rotation. However, in the plane perpendicular to this (i.e., the plane through the poles of the spinning system) contraction would continue. Thus we would end up with a disk-shaped galaxy.

We can now see an explanation of the great differences between the disk-shaped spiral galaxies and the more globular elliptic galaxies. A protogalaxy which started with considerable rotation (high angular momentum) would spin out to a disk shape and, because its contraction in the plane of rotation was severely limited, would never become very dense and would be left with a great deal of uncondensed gas, especially in its outer parts. On the other hand, a system that started with little rotational momentum could contract far more in the plane of rotation; thus it could become extremely dense and ultimately might condense almost completely into stars, forming a smooth, dense ball—i.e., an elliptical galaxy.

We can also understand the two-population structure of our own galaxy. The Population II stars and globular clusters of stars must have been born in the earliest stages of the protogalaxy's evolution; they may even have been in existence before the protogalaxy detached itself as a separate system. This swarm of stars could not contract to a disk. In the course of time the originally chaotic swarm became a regular, spherical group, concentrated toward the center. In the disk part of our galaxy, on the other hand, stars condensed only after the disk itself had been formed by the contraction and flattening of the gas clouds. The earliest stars formed in the disk are now so old that we identify them as of the Population II type (disk Population II). In the huge, more or less homogeneous disk of gas that remained, the younger

Population I stars were formed as time went on and are still being born at the present time.

We are still left with the problem of accounting for the spiral arms. Some have suggested that the arms developed from close encounters between galaxies, which drew out the long arms by tidal force. However, such encounters are too rare now to account for all the spiral systems we see, and even if we assume that they occurred frequently in an early stage of the universe, there are serious objections to this theory. One thing is clear: the spiral structure depends on some property of the interstellar gas—without clouds of gas there can be no spiral arms. This has been interestingly confirmed by the observation of Baade and Lyman Spitzer, Jr., that spiral systems are rare in dense clusters of galaxies. Presumably the galaxies in such clusters interpenetrate one another occasionally, and in so doing remove the gas (see following chapter).

The answer to the enigma of the spirals probably will come from more thorough study of the spiral arms in our own galaxy, which have just begun to be explored. They have already presented us with some new and difficult questions. One of the most challenging has to do with the fact that at a point halfway between us and the center of the galaxy the arms are rotating around the center with about twice the speed of their rotation in our neighborhood, out near the galaxy's rim. This differential rotation should be winding the arms more and more tightly around the core, and it can be calculated that under such conditions spiral arms could not survive for more than a few hundred million years. In view of the fact that a majority of the galaxies we observe are spirals, the spiral structure must have a longer duration than this small fraction of the lifetime of a galaxy. We can only suppose that some compensating process is at work building up the arms while the rotation pulls them in. Perhaps one side of an arm collects material while the other side evaporates it, so that the arm holds a constant position. Calculations indicate that the arms may be maintained in this way by the same

mechanism thought to be responsible for the internal motions in interstellar gas.

The new view of the origin of the galaxies gives us, among other things, a new perspective on the conditions existing in the universe when the galaxies were first formed. It should enable us to deduce, from the present nature and behavior of the galaxies, something about the density and motions of the matter in the universe at that far-off epoch.

COLLIDING GALAXIES

by Rudolph Minkowski

During the past ten years radio astronomers have discovered a new kind of object in the sky. These objects are distinct sources of radio energy, just as stars or galaxies are discrete sources of light. Some of the radio sources can be identified with well-known visible objects, such as the spectacular Crab Nebula, which marks the location of an exploded star in our galaxy. But curiously the two most powerful radio sources are located at points in the sky where no conspicuous objects appear.

One of these lies in the constellation Cygnus. As the resolving power of radio telescopes was improved, the search for the radio source narrowed down to an area containing a faint object far out beyond the stars of our own galaxy. In 1951 F. G. Smith of the University of Cambridge, using a British radio interferometer of high resolving power (see the chapter on radio galaxies, page 112), boxed the source of the strong radio emission, called Cygnus A, within an area of less than one square minute of arc.

Walter Baade thereupon took a picture of the object with the 200-inch telescope on Palomar Mountain, and made a surprising discovery. What had appeared in plates made with smaller telescopes as a single galaxy turned out to be an unusual system composed of two galaxies, one superposed on the other. Furthermore, it became plain that the two are not separate galaxies in the same line of sight but are actually in close contact, for their nuclei are strongly distorted by gravitational interaction. The centers of the two systems are only about two seconds of arc apart; evidently the galaxies have penetrated far into each other. In short, there was every indication that in Cygnus A we are beholding a close collision between two galaxies (Plate 19).

This was soon confirmed by a beautiful proof. What should we expect to see when galaxies collide? It is altogether unlikely that their individual stars will collide, because the average distance between stars in a galaxy is immense. In an encounter between two galaxies the gravitational attraction of their stars will perturb the stars' motions and distort the structure of the galaxies, but the stars themselves would not be expected to undergo any observable change. We should, however, expect to see an important effect on clouds of dust and gas in the galaxies. The gas and dust particles *will* collide. At velocities in the range of hundreds to thousands of miles per second such collisions should heat the gas to temperatures of one million to 100 million degrees.

Assuming that Cygnus A is a galactic collision, then, its gas should be very hot, and the atoms should be in a highly excited state. There is a way to test whether this is so: namely, examine the spectrum of light from the hot material. We can read the degree of excitation of atoms in a glowing gas from the character of the lines in its spectrum. Studies of spectra of Cygnus A made with the 100-inch and 200-inch telescopes fully confirmed the reality of the collision. Almost half of the light of the system is in broadened lines of hydrogen and in "forbidden" lines of emission by oxygen, neon, sulfur and iron in highly ionized states.

The strong emission lines of this spectrum, incidentally, made it possible to measure the red shift of Cygnus A, and thereby its distance from us, with great accuracy. The pair of colliding galaxies is 270 million light-years away.

The case of the colliding galaxies in Cygnus has great interest in itself, but it has even more interest for cosmology. It has already yielded a discovery of first importance. A. E. Lilley and E. F. McClain of the Naval Research Laboratory found that the 21-centimeter radio line of hydrogen in Cygnus A shows a Doppler shift precisely in accord with the red shift of its light. This seems to verify that the red shift is a Doppler effect and therefore a measure of the velocity and distance of galaxies.

The measurement of a radio "red shift" is particularly exciting

because radio astronomy promises a long extension of our reach into space. The radio signal from Cygnus A could be detected by our present radio telescopes even if it were only one 3,000th as intense as it is; thus we could detect such collisions at vastly greater distances—distances far beyond the range of the 200-inch optical telescope. All in all, there is solid ground for expecting that radio observation of colliding galaxies will answer many cosmological questions which optical astronomy cannot.

How frequent are collisions between galaxies? That depends in part on how we define a "collision." In the case of Cygnus A the two galaxies have penetrated far into each other: their central bodies, or nuclei, are only about 3,000 light-years apart. Such an event must be exceedingly rare. If we take a collision to mean any encounter in which the galaxies approach close enough to touch at the outer boundaries of their thin reaches of matter, the probability (i.e., estimated frequency) of collisions becomes much greater. Assuming that the radius of a typical galaxy is 15,000 light-years, two such galaxies will be in collision, by our definition, whenever their centers are no more than 30,000 light-years apart. In the average region of space the average distance between galaxies is about three million light-years. On the basis of these figures and of what we know about the motions of galaxies, we can calculate that in the volume of space out to 250 million light-years from us, which is estimated to contain about two billion galaxies, some 10 collisions would be in progress at the present time if the galaxies were evenly distributed in space.

Of course this theoretical average must be adjusted to take account of the actual sizes and distributions of galaxies. Our own Milky Way has a radius of about 40,000 light-years; some spiral galaxies have arms extending out to 60,000 light-years. What is much more important, many galaxies are grouped in clusters, where the average distance of separation is much less than three million light-years. For example, the central part of the Coma cluster of galaxies has some 500 galaxies concentrated within a space about 2.6 million light-years across. Here, we can estimate, at least two collisions should be in progress at this

moment; the number is probably higher because the cluster increases in density toward its center. Cygnus A lies in another fairly rich cluster of galaxies. Collisions of its type, however, must still be rare, for the calculations indicate that a collision as close as this one is 1,000 times less frequent than the slightly overlapping contacts we have been discussing. Thus we should not expect to find more than a few clusters containing objects like Cygnus A.

Even before an actual instance of collision was found, Baade and Lyman Spitzer, Jr., of Princeton University had calculated the high probability of collisions in clusters and had assigned to the process an important function in the evolution of galaxies. Most of the galaxies in a condensed cluster such as Coma (Plate 22) have a general form characteristic of the spiral type of galaxy (notably a flattened disk shape) but lack the arms and clouds of gas typical of the spiral systems. Spitzer and Baade proposed that these cluster galaxies were originally spirals and that collisions had swept them clear of their clouds of gas and dust, in which the hot stars characteristic of spiral arms are formed.

A galaxy would have to undergo a fairly large number of collisions to lose all its interstellar matter. It can be shown, however, that a sufficient number of collisions is likely to occur in a dense cluster. In the Coma cluster, for example, a galaxy moving roughly along the radius of the system and through the region of its center would have suffered between five and 30 collisions (the number depending on the size of the galaxy) during the five-billion-year lifetime of the universe. This order of probability is adequate to account for the removal of interstellar matter from most of the members of the Coma cluster. Clouds of material removed from a spiral galaxy may assist in sweeping dust and gas out of other spirals in the cluster. David S. Heeschen of the Harvard College Observatory has in fact found evidence—21-centimeter radiation—that the Coma cluster contains a large amount of hydrogen gas.

In a loose cluster, where collisions are less frequent, we should

expect to find that many of the galaxies still possess spiral arms and gas clouds. There is a loose cluster in Virgo which actually shows such a picture.

The discovery of the Cygnus A collision naturally inspired astronomers to an intensive search for collisions in other radio "stars." A few have been found, though none as intense as Cygnus A. It has been possible to study one in considerable detail. The object is known as NGC 1275, the brightest member of a cluster of visible galaxies in the constellation Perseus. The light spectrum of this object has long been known to be peculiar, with the emission lines that indicate high excitation. Now a detailed spectroscopic study of the object has shown clearly that it consists of two galaxies, colliding with a relative velocity of about 2,000 miles per second. One of the pair is a tightly wound spiral; the other, a very loose spiral with grossly distorted arms. The two galaxies are inclined at a small angle to each other, and we see both almost face on. The loose system is tilted to our line of sight in a direction such that the top or northerly part of the galaxy is nearer to us and the southerly part is away. In the upper portion of the overlapping image of the two galaxies we can see two distinct spectra of relatively low excitation. These evidently come from separate masses of gas lying respectively in the loose spiral and in the compact spiral behind it. This must mean that these sections of the two galaxies have not yet collided. But at a line just above the nucleus of the compact galaxy there is an abrupt change in the spectrum: here we get only one spectrum of emission, from gas in a very excited state. The line apparently marks the section where the loose spiral galaxy has entered the compact one and is slicing through it. Below this line we see but a single spectrum, which indicates that the forward edge of the loose galaxy has passed through the lower part of the compact system and left it in an agitated state. As the loose galaxy continues to move through, the line of collision should gradually shift upward, or northward.

We now have a fair picture of the sequence of conditions that leads to radio emission in collisions. In general it is clear that strong radio emission occurs only when the galaxies actually penetrate each other. But we still have little idea as to the mechanism responsible for generating radio energy of such strength. Astrophysicists have been hunting for such a mechanism, but so far with little success. The intensity and the spectrum of the radio emission show that the radio energy cannot arise merely from the heating of the gas that occurs as a result of the collision.

In Cygnus A the strong radio signals seem to be coming from two outlying parts of the system separated by about 120,000 light-years, almost three times the visible extent of the system. We know that galaxies extend far into space beyond their bright central masses, and we know that a collision may scatter their gas clouds widely. But we cannot explain how the faint outer regions can emit radio energy at the huge rate of 10^{44} ergs per second—10 times the output of light by the colliding galaxies. Energy of this order can be provided by the relative kinetic energy of the colliding systems, but this energy is concentrated in their central masses.

It is tempting to suppose that at least part of the radio energy in collisions may be produced by a process already observed in the Crab Nebula, whose radio emission is generated by the motion of high-energy electrons in a magnetic field. We cannot, however, exclude the possibility of other processes. We know, for instance, that our sun generates radio energy by a mechanism different from that of the Crab Nebula.

The colliding galaxies thus offer many interesting problems: how galaxies interact and generate radio energy during collisions, how frequently collisions occur, and so on. The answers to these questions should lead to a far better understanding of our universe.

PART 4 COSMOLOGICAL THEORY

I. THE EVOLUTIONARY UNIVERSE

by George Gamow

A many-sided scientist whose interests have led him into a variety of fields, George Gamow is now professor of physics at the University of Colorado. A quixotic streak, which has enlivened his many popular books and articles, sometimes extends to his most serious scientific publications. On one occasion when he and Ralph A. Alpher were preparing a paper, they invited Hans Bethe of Cornell University to collaborate with them. The paper, which happened to concern the beginning of the universe, was therefore most appropriately authored by Alpher, Bethe and Gamow. He was born in Odessa and educated at the University of Leningrad. Before settling in the United States in 1934, he taught and did research at the universities of Göttingen, Copenhagen and Cambridge. He has worked in many fields: biology at the molecular level, the theory of radioactivity, the structure of the atomic nucleus, thermonuclear reactions in stars, the origin of the chemical elements and, of course, the evolution of the universe.

II. THE STEADY-STATE UNIVERSE

by Fred Hoyle

The author is lecturer in mathematics at St. John's College, University of Cambridge. He was born in Yorkshire in 1914; by the time he was six, he had taught himself the multiplication tables up to 12 times 12. Not all of his early experience with mathematics was so prodigious. One of his first teachers slapped him for miscounting the number of petals on a flower; his sense of justice was so outraged that he refused to return to the school. When he was 13 his parents bought him a three-inch telescope and allowed him to sit up all night with it. In 1930 he

won a prize fellowship at St. John's. The next year he joined an Admiralty research group, but continued to devote much of his spare time to astronomy. After the war he returned to Cambridge.

THE EVOLUTIONARY UNIVERSE

by George Gamow

COSMOLOGY is the study of the general nature of the universe in space and in time—what it is now, what it was in the past and what it is likely to be in the future. Since the only forces at work between the galaxies that make up the material universe are the forces of gravity, the cosmological problem is closely connected with the theory of gravitation, in particular with its modern version as comprised in Albert Einstein's general theory of relativity. In the frame of this theory the properties of space, time and gravitation are merged into one harmonious and elegant picture.

The basic cosmological notion of general relativity grew out of the work of great mathematicians of the nineteenth century. In the middle of the last century two inquisitive mathematical minds—a Russian named Nikolai Lobachevski and a Hungarian named János Bolyai—discovered that the classical geometry of Euclid was not the only possible geometry: in fact, they succeeded in constructing a geometry which was fully as logical and self-consistent as the Euclidean. They began by overthrowing Euclid's axiom about parallel lines: namely, that only one parallel to a given straight line can be drawn through a point not on that line. Lobachevski and Bolyai both conceived a system of geometry in which a great number of lines parallel to a given line could be drawn through a point outside the line.

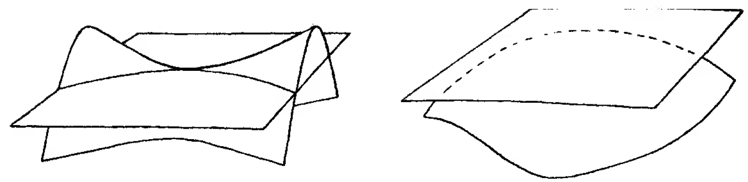
To illustrate the differences between Euclidean geometry and their non-Euclidean system it is simplest to consider just two dimensions—that is, the geometry of surfaces. In our schoolbooks this is known as "plane geometry," because the Euclidean surface is a flat surface. Suppose, now, we examine the properties of a

two-dimensional geometry constructed not on a plane surface but on a curved surface. For the system of Lobachevski and Bolyai we must take the curvature of the surface to be "negative," which means that the curvature is not like that of the surface of a sphere but like that of a saddle (see diagrams opposite). Now if we are to draw parallel lines or any figure (e.g., a triangle) on this surface, we must decide first of all how we shall define a "straight line," equivalent to the straight line of plane geometry. The most reasonable definition of a straight line in Euclidean geometry is that it is the path of the shortest distance between two points. On a curved surface the line, so defined, becomes a curved line known as a "geodesic."

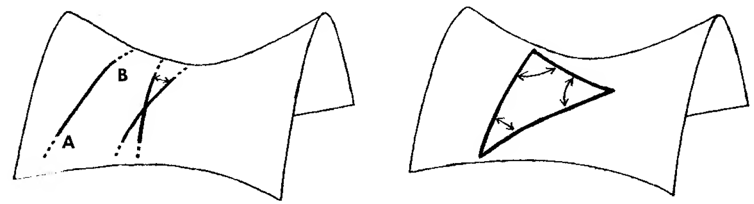
Considering a surface curved like a saddle, we find that, given a "straight" line or geodesic, we can draw through a point outside that line a great many geodesics which will never intersect the given line, no matter how far they are extended. They are therefore parallel to it, by the definition of parallel. The possible parallels to the line fall within certain limits, indicated by the intersecting lines in the middle drawing.

As a consequence of the overthrow of Euclid's axiom on parallel lines, many of his theorems are demolished in the new geometry. For example, the Euclidean theorem that the sum of the three angles of a triangle is 180 degrees no longer holds on a curved surface. On the saddle-shaped surface the angles of a triangle formed by three geodesics always add up to less than 180 degrees, the actual sum depending on the size of the triangle. Further, a circle on the saddle surface does not have the same properties as a circle in plane geometry. On a flat surface the circumference of a circle increases in proportion to the increase in diameter, and the area of a circle increases in proportion to the square of the increase in diameter. But on a saddle surface both the circumference and the area of a circle increase at *faster* rates than on a flat surface with increasing diameter.

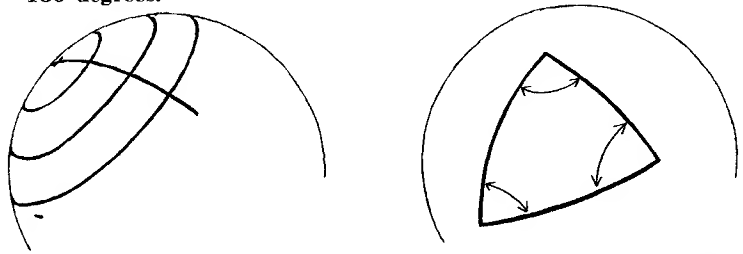
After Lobachevski and Bolyai, the German mathematician Bernhard Riemann constructed another non-Euclidean geometry



Negative and positive curvature of space is suggested by this two-dimensional analogy. The saddle-shaped surface at left, which lies on both sides of a tangential plane, is negatively curved. The spherical surface at right, which lies on one side of a tangential plane, is positively curved. If space is negatively curved, the universe is infinite; if it is positively curved, the universe is finite.



On a negatively curved surface the shortest distance between two points is not a straight line but a curved "geodesic," such as the line AB on the left. On a plane surface only one parallel to a given straight line can be drawn through a point not on that line; on a negatively curved surface many geodesics can be drawn through a point not on a given geodesic without ever intersecting it. These "parallel" lines will fall within the limits indicated by the arrow between the intersecting lines at left. On a plane surface the angles of a triangle add up to 180 degrees; on the negatively curved surface at the right, they add up to less than 180 degrees.



On a positively curved surface the shortest distance between two points follows a great circle, a closed line passing through opposite points on the surface (*single curved line at left*). In this geometry there are no parallel lines because any two great circles must intersect. The circumference of a circle increases more slowly with diameter than on a flat surface, and the area similarly increases more slowly (*concentric circles at left*). The angles of a triangle on the surface (*right*) add up to more than 180 degrees.

whose two-dimensional model is a surface of positive, rather than negative, curvature—that is, the surface of a sphere. In this case a geodesic line is simply a great circle around the sphere or a segment of such a circle, and since any two great circles must intersect at two points (the poles), there are no parallel lines at all in this geometry. Again the sum of the three angles of a triangle is not 180 degrees: in this case it is always *more* than 180. The circumference of a circle now increases at a rate *slower* than in proportion to its increase in diameter, and its area increases more slowly than the square of the diameter.

Now all this is not merely an exercise in abstract reasoning but bears directly on the geometry of the universe in which we live. Is the space of our universe "flat," as Euclid assumed, or is it curved negatively (per Lobachevski and Bolyai) or curved positively (Riemann)? If we were two-dimensional creatures living in a two-dimensional universe, we could tell whether we were living on a flat or a curved surface by studying the properties of triangles and circles drawn on that surface. Similarly, as three-dimensional beings living in three-dimensional space we should be able, by studying geometrical properties of that space, to decide what the curvature of our space is. Riemann in fact developed mathematical formulas describing the properties of various kinds of curved space in three and more dimensions. In the early years of this century Einstein conceived the idea of the universe as a curved system in four dimensions, embodying time as the fourth dimension, and he proceeded to apply Riemann's formulas to test his idea.

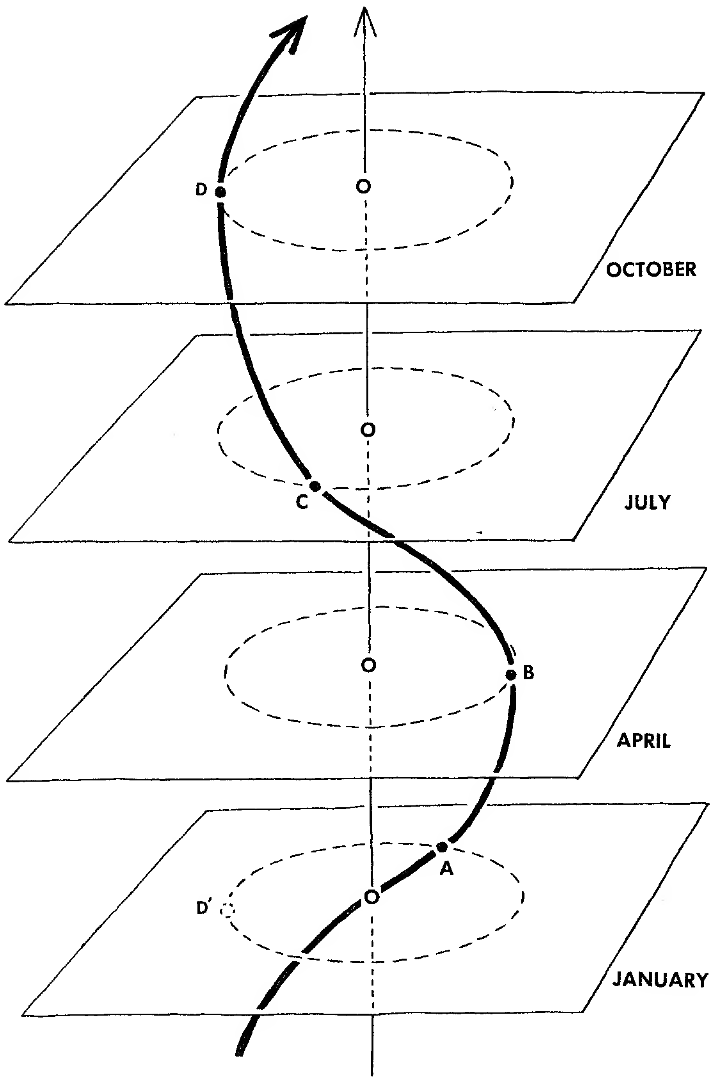
Einstein showed that time can be considered a fourth co-ordinate supplementing the three co-ordinates of space. He connected space and time, thus establishing a "space-time continuum," by means of the speed of light as a link between time and space dimensions. However, recognizing that space and time are physically different entities, he employed the imaginary number $\sqrt{-1}$, or i, to express the unit of time mathematically and

make the time co-ordinate formally equivalent to the three co-ordinates of space.

In his special theory of relativity Einstein made the geometry of the time-space continuum strictly Euclidean, that is, flat. The great idea that he introduced later in his general theory was that gravitation, whose effects had been neglected in the special theory, must make it curved. He saw that the gravitational effect of the masses distributed in space and moving in time was equivalent to curvature of the four-dimensional space-time continuum. In place of the classical Newtonian statement that "the sun produces a field of forces which impels the earth to deviate from straight-line motion and to move in a circle around the sun," Einstein substituted a statement to the effect that "the presence of the sun causes a curvature of the space-time continuum in its neighborhood."

The motion of an object in the space-time continuum can be represented by a curve called the object's "world line." For example, the world line of the earth's travel around the sun in time is pictured in the drawing on the next page. (Space must be represented here in only two dimensions; it would be impossible for a three-dimensional artist to draw the fourth dimension in this scheme, but since the orbit of the earth around the sun lies in a single plane, the omission is unimportant.) Einstein declared, in effect: "The world line of the earth is a geodesic in the curved four-dimensional space around the sun." In other words, the line ABCD in the drawing corresponds to the shortest *four-dimensional* distance between the position of the earth in January (at A) and its position in October (at D).

Einstein's idea of the gravitational curvature of space-time was, of course, triumphantly affirmed by the discovery that it could account for the previously inexplicable perturbations in the motion of Mercury at its closest approach to the sun and the discovery that light rays are deflected by the sun's gravitational field. Einstein next attempted to apply the idea to the universe as a whole. Does it have a general curvature, similar to the local



Motion of body in the curved "space-time continuum" of Albert Einstein is represented by the "world line" of the earth's motion around the sun. Here the sun is the small open circle in each of the four planes. The earth is the black dot on the elliptical orbit. Each plane shows the position of the earth at a month of the year. The world line connects points A, B, C, D.

curvature in the sun's gravitational field? He now had to consider not a single center of gravitational force but countless centers of attraction in a universe full of matter concentrated in galaxies whose distribution fluctuates considerably from region to region in space. However, in the large-scale view the galaxies are spread fairly uniformly throughout space as far out as our biggest telescopes can see, and we can justifiably "smooth out" its matter to a general average (which comes to about one hydrogen atom per cubic meter). On this assumption the universe as a whole has a smooth general curvature.

But if the space of the universe is curved, what is the sign of this curvature? Is it positive, as in our two-dimensional analogy of the surface of a sphere, or is it negative, as in the case of a saddle surface? And, since we cannot consider space alone, how is this space curvature related to time?

Analyzing the pertinent mathematical equations, Einstein came to the conclusion that the curvature of space must be independent of time, i.e., that the universe as a whole must be unchanging (though it changes internally). However, he found to his surprise that there was no solution of the equations that would permit a static cosmos. To repair the situation, Einstein was forced to introduce an additional hypothesis which amounted to the assumption that a new kind of force was acting among the galaxies. This hypothetical force had to be independent of mass (being the same for an apple, the moon and the sun!) and to gain in strength with increasing distance between the interacting objects (as no other forces ever do in physics!).

Einstein's new force, called "cosmic repulsion," allowed two mathematical models of a static universe. One solution, which was worked out by Einstein himself and became known as "Einstein's spherical universe," gave the space of the cosmos a positive curvature. Like a sphere, this universe was closed and thus had a finite volume. The space co-ordinates in Einstein's spherical universe were curved in the same way as the latitude or longitude coordinates on the surface of the earth. However, the time axis of

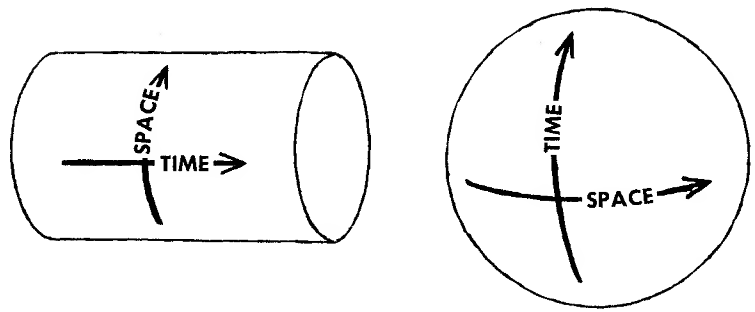
the space-time continuum ran quite straight, as in the good old classical physics. This means that no cosmic event would ever recur. The two-dimensional analogy of Einstein's space-time continuum is the surface of a cylinder, with the time axis running parallel to the axis of the cylinder and the space axis perpendicular to it (see drawing at left on opposite page).

The other static solution based on the mysterious repulsion forces was discovered by the Dutch mathematician Willem de Sitter. In his model of the universe both space and time were curved. Its geometry was similar to that of a globe, with longitude serving as the space co-ordinate and latitude as time (drawing at right on opposite page).

Unhappily astronomical observations contradicted both Einstein's and de Sitter's static models of the universe, and they were soon abandoned.

In the year 1922 a major turning point came in the cosmological problem. A Russian mathematician, Alexander A. Friedman (from whom the author of this chapter learned his relativity), discovered an error in Einstein's proof for a static universe. In carrying out his proof Einstein had divided both sides of an equation by a quantity which, Friedman found, could become zero under certain circumstances. Since division by zero is not permitted in algebraic computations, the possibility of a nonstatic universe could not be excluded under the circumstances in question. Friedman showed that two nonstatic models were possible. One pictured the universe as expanding with time; the other, contracting.

Einstein quickly recognized the importance of this discovery. In the last edition of his book *The Meaning of Relativity* he wrote: "The mathematician Friedman found a way out of this dilemma. He showed that it is possible, according to the field equations, to have a finite density in the whole (three-dimensional) space, without enlarging these field equations ad hoc." Einstein remarked to me many years ago that the cosmic repul-



Spherical universe of Einstein may be represented in two dimensions by a cylinder (*left*). Its space co-ordinates were positively curved but its time co-ordinate was straight. The spherical universe of Willem de Sitter had positively curved co-ordinates (*right*).

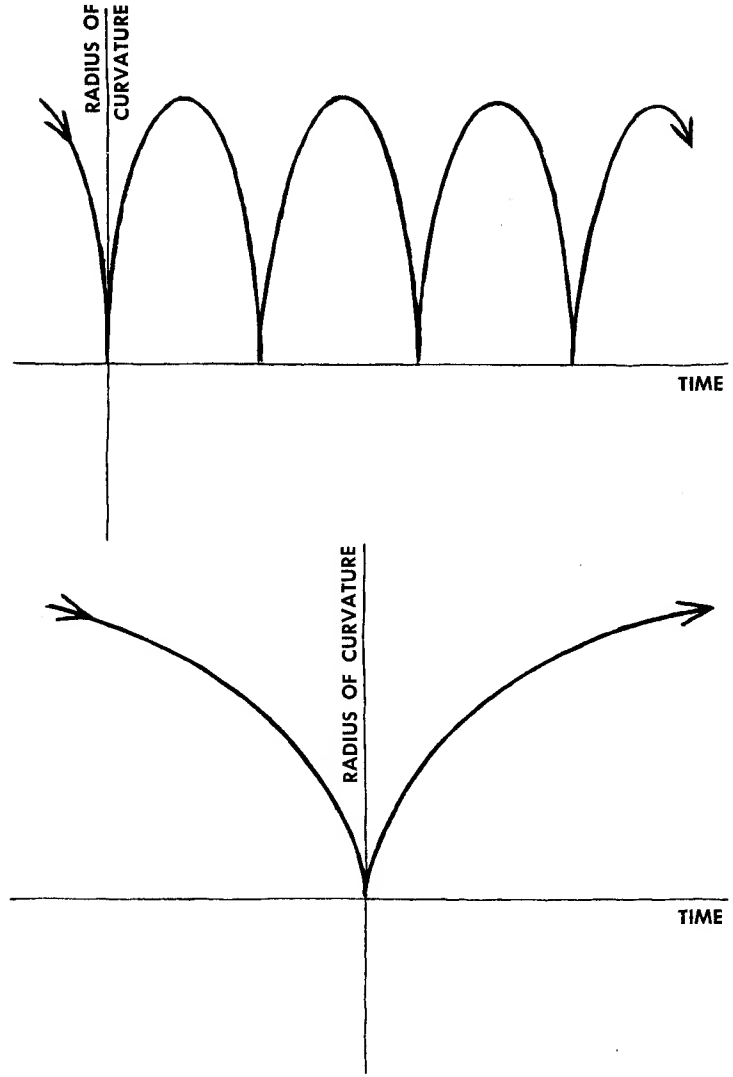
sion idea was the biggest blunder he had made in his entire life.

Almost at the very moment that Friedman was discovering the possibility of an expanding universe by mathematical reasoning, Edwin P. Hubble at the Mount Wilson Observatory on the other side of the world found the first evidence of actual physical expansion through his telescope. He made a compilation of the distances of a number of far galaxies, whose light was shifted toward the red end of the spectrum, and it was soon found that the extent of the shift was in direct proportion to a galaxy's distance from us, as estimated by its faintness. Hubble and others interpreted the red shift as the Doppler effect—the well-known phenomenon of lengthening of wave lengths from any radiating source that is moving rapidly away (a train whistle, a source of light or whatever). To date there has been no other reasonable explanation of the galaxies' red shift. If the explanation is correct, it means that the galaxies are all moving away from one another with increasing velocity as they move farther apart.

Thus Friedman and Hubble laid the foundation for the theory of the expanding universe. The theory was soon developed further by a Belgian theoretical astronomer, Georges Lemaître. He proposed that our universe started from a highly compressed and extremely hot state which he called the "primeval atom." (Modern physicists would prefer the term "primeval nucleus.") As this matter expanded, it gradually thinned out, cooled down and reaggregated in stars and galaxies, giving rise to the highly complex structure of the universe as we know it today.

Until a few years ago the theory of the expanding universe lay under the cloud of a very serious contradiction. The measurements of the speed of flight of the galaxies and their distances from us indicated that the expansion had started about 1.8 billion years ago. On the other hand, measurements of the age of ancient rocks in the earth by the clock of radioactivity (i.e., the decay of uranium to lead) showed that some of the rocks were at least three billion years old; more recent estimates based on other radioactive elements raise the age of the earth's crust to almost five billion years. Clearly a universe 1.8 billion years old could not contain five-billion-year-old rocks! Happily the contradiction has now been disposed of by Walter Baade's recent discovery that the distance yardstick (based on the periods of variable stars) was faulty and that the distances between galaxies are more than twice as great as they were thought to be. This change in distances raises the age of the universe to five billion years or more.

Friedman's solution of Einstein's cosmological equation, as I mentioned, permits two kinds of universe. We can call one the "pulsating" universe. This model says that when the universe has reached a certain maximum permissible expansion it will begin to contract; that it will shrink until its matter has been compressed to a certain maximum density, possibly that of atomic nuclear material, which is a hundred million million times denser than water; that it will then begin to expand again—and so on through the cycle ad infinitum. The other model is a "hyperbolic" one: it suggests that from an infinitely thin state an eternity ago



Pulsating and hyperbolic universes are represented by curves. The pulsating universe at the top repeatedly expands to a maximum permissible density and contracts to a minimum permissible density. The hyperbolic universe at the bottom contracts and then expands indefinitely.

the universe contracted until it reached the maximum density, from which it rebounded to an unlimited expansion which will go on indefinitely in the future.

The question whether our universe is actually "pulsating" or "hyperbolic" should be decidable from the present rate of its expansion. The situation is analogous to the case of a rocket shot from the surface of the earth. If the velocity of the rocket is less than seven miles per second—the "escape velocity"—the rocket will climb only to a certain height and then fall back to the earth. (If it were completely elastic, it would bounce up again, etc., etc.) On the other hand, a rocket shot with a velocity of more than seven miles per second will escape from the earth's gravitational field and disappear in space. The case of the receding system of galaxies is very similar to that of an escape rocket, except that instead of just two interacting bodies (the rocket and the earth) we have an unlimited number of them escaping from one another. We find that the galaxies are fleeing from one another at seven times the velocity necessary for mutual escape.

Thus we may conclude that our universe corresponds to the "hyperbolic" model, so that its present expansion will never stop. We must make one reservation. The estimate of the necessary escape velocity is based on the assumption that practically all the mass of the universe is concentrated in galaxies. If intergalactic space contained matter whose total mass was more than seven times that in the galaxies, we would have to reverse our conclusion and decide that the universe is pulsating. There has been no indication so far, however, that any matter exists in intergalactic space, and it could have escaped detection only if it were in the form of pure hydrogen gas, without other gases or dust.

Is the universe finite or infinite This resolves itself into the question: Is the curvature of space positive or negative—closed like that of a sphere, or open like that of a saddle? We can look for the answer by studying the geometrical properties of its three-dimensional space, just as we examined the properties of

figures on two-dimensional surfaces. The most convenient property to investigate astronomically is the relation between the volume of a sphere and its radius.

We saw that, in the two-dimensional case, the area of a circle increases with increasing radius at a faster rate on a negatively curved surface than on a Euclidean or flat surface; and that on a positively curved surface the relative rate of increase is slower. Similarly the increase of volume is faster in negatively curved space, slower in positively curved space. In Euclidean space the volume of a sphere would increase in proportion to the cube, or third power of the increase in radius. In negatively curved space the volume would increase faster than this; in positively curved space, slower. Thus if we look into space and find that the volume of successively larger spheres, as measured by a count of the galaxies within them, increases faster than the cube of the distance to the limit of the sphere (the radius), we can conclude that the space of our universe has negative curvature, and therefore is open and infinite. By the same token, if the number of galaxies increases at a rate slower than the cube of the distance, we live in a universe of positive curvature—closed and finite.

Following this idea, Hubble undertook to study the increase in number of galaxies with distance. He estimated the distances of the remote galaxies by their relative faintness: galaxies vary considerably in intrinsic brightness, but over a very large number of galaxies these variations are expected to average out. Hubble's calculations produced the conclusion that the universe is a closed system—a small universe only a few billion light-years in radius!

We know now that the scale he was using was wrong: with the new yardstick the universe would be more than twice as large as he calculated. But there is a more fundamental doubt about his result. The whole method is based on the assumption that the intrinsic brightness of a galaxy remains constant. What if it changes with time? We are seeing the light of the distant galaxies as it was emitted at widely different times in the past—500 million, a billion, two billion years ago. If the stars in the galaxies are

burning out, the galaxies must dim as they grow older. A galaxy two billion light-years away cannot be put on the same distance scale with a galaxy 500 million light-years away unless we take into account the fact that we are seeing the nearer galaxy at an older, and less bright, age. The remote galaxy is farther away than a mere comparison of the luminosity of the two would suggest.

When a correction is made for the assumed decline in brightness with age, the more distant galaxies are spread out to farther distances than Hubble assumed. In fact, the calculations of volume are changed so drastically that we may have to reverse the conclusion about the curvature of space. We are not sure, because we do not yet know enough about the evolution of galaxies. But if we find that galaxies wane in intrinsic brightness by only a few per cent in a billion years, we shall have to conclude that space is curved negatively and the universe is infinite.

Actually there is another line of reasoning which supports the side of infinity. Our universe seems to be hyperbolic and ever-expanding. Mathematical solutions of fundamental cosmological equations indicate that such a universe is open and infinite.

We have reviewed the questions that dominated the thinking of cosmologists during the first half of this century: the conception of a four-dimensional space-time continuum, of curved space, of an expanding universe and of a cosmos which is either finite or infinite. Now we must consider the major present issue in cosmology: Is the universe in truth evolving, or is it in a steady state of equilibrium which has always existed and will go on through eternity? Most cosmologists take the evolutionary view. But in 1951 a group at the University of Cambridge, whose chief spokesman has been Fred Hoyle, advanced the steady-state idea. Essentially their theory is that the universe is infinite in space and time, that it has neither a beginning nor an end, that the density of its matter remains constant, that new matter is

steadily being created in space at a rate which exactly compensates for the thinning of matter by expansion, that as a consequence new galaxies are continually being born, and that the galaxies of the universe therefore range in age from mere youngsters to veterans of 5, 10, 20 and more billions of years. In my opinion this theory must be considered very questionable because of the simple fact (apart from other reasons) that the galaxies in our neighborhood all seem to be of the same age as our own Milky Way. But the issue is many-sided and fundamental, and can be settled only by extended study of the universe as far as we can observe it. Hoyle presents the steady-state view in the following chapter. Here I shall summarize the evolutionary theory.

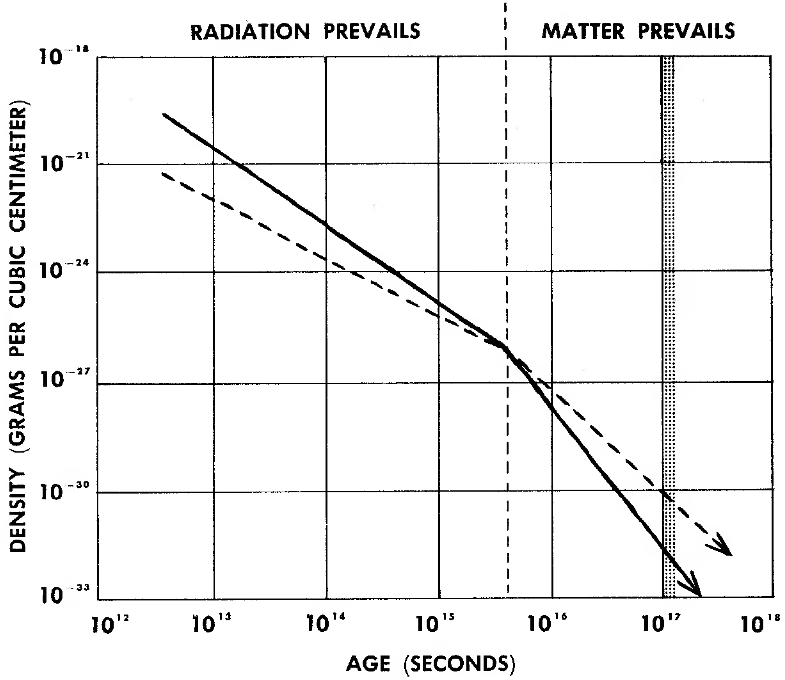
We assume that the universe started from a very dense state of matter. In the early stages of its expansion, radiant energy was dominant over the mass of matter. We can measure energy and matter on a common scale by means of the well-known equation $E = mc^2$, which says that the energy equivalent of matter is the mass of the matter multiplied by the square of the velocity of light. Energy can be translated into mass, conversely, by dividing the energy quantity by c^2. Thus we can speak of the "mass density" of energy. Now, at the beginning the mass density of the radiant energy was incomparably greater than the density of the matter in the universe. But in an expanding system the density of radiant energy decreases faster than does the density of matter. The former thins out as the fourth power of the distance of expansion: as the radius of the system doubles, the density of radiant energy drops to one sixteenth. The density of matter declines as the third power; a doubling of the radius means an eightfold increase in volume, or eightfold decrease in density.

Assuming that the universe at the beginning was under absolute rule by radiant energy, we can calculate that the temperature of the universe was 250 million degrees when it was one hour old, dropped to 6,000 degrees (the present temperature of our sun's surface) when it was 200,000 years old and had fallen to about

100 degrees below the freezing point of water when the universe reached its 250-millionth birthday.

This particular birthday was a crucial one in the life of the universe. It was the point at which the density of ordinary matter became greater than the mass density of radiant energy, because of the more rapid fall of the latter (see chart opposite). The switch from the reign of radiation to the reign of matter profoundly changed matter's behavior. During the eons of its subjugation to the will of radiant energy (i.e., light), it must have been spread uniformly through space in the form of thin gas. But as soon as matter became gravitationally more important than the radiant energy, it began to acquire a more interesting character. James Jeans, in his classic studies of the physics of such a situation, proved half a century ago that a gravitating gas filling a very large volume is bound to break up into individual "gas balls," the size of which is determined by the density and the temperature of the gas. Thus in the year 250,000,000 A.B.E. (after the beginning of expansion), when matter was freed from the dictatorship of radiant energy, the gas broke up into giant gas clouds, slowly drifting apart as the universe continued to expand. Applying Jeans's mathematical formula for the process to the gas filling the universe at that time, I have found that these primordial balls of gas would have had just about the mass that the galaxies of stars possess today. They were then only "protogalaxies"—cold, dark and chaotic. But their gas soon condensed into stars and formed the galaxies as we see them now.

A central question in this picture of the evolutionary universe is the problem of accounting for the formation of the varied kinds of matter composing it—i.e., the chemical elements. The question is discussed in detail in an earlier chapter (see page 17). My belief is that at the start matter was composed simply of protons, neutrons and electrons. After five minutes the universe must have cooled enough to permit the aggregation of protons and neutrons into larger units, from deuterons (one neutron and



Relative density of matter and radiation is reversed during the history of an evolutionary universe. Up to 250 million years (*broken vertical line*) the mass density of radiation (*solid curve*) is greater than that of matter (*broken curve*). After that the density of matter is greater, permitting the formation of huge gas clouds. The vertical gray line indicates the present.

one proton) up to the heaviest elements. This process must have ended after about 30 minutes, for by that time the temperature of the expanding universe must have dropped below the threshold of thermonuclear reactions among light elements, and the neutrons must have been used up in element-building or been converted to protons.

To many a reader the statement that the present chemical constitution of our universe was decided in half an hour five billion years ago will sound nonsensical. But consider a spot of ground on the atomic proving ground in Nevada where an atomic bomb was exploded in 1953. Within one microsecond the nuclear reactions generated by the bomb produced a variety of fission

products. In 1956, 100 million million microseconds later, the site was still "hot" with the surviving fission products. The ratio of one microsecond to three years is the same as the ratio of half an hour to five billion years! If we can accept a time ratio of this order in the one case, why not in the other?

The late Enrico Fermi and Anthony L. Turkevich at the Institute for Nuclear Studies of the University of Chicago undertook a detailed study of thermonuclear reactions such as must have taken place during the first half hour of the universe's expansion. They concluded that the reactions would have produced about equal amounts of hydrogen and helium, making up 99 per cent of the total material, and about 1 per cent of deuterium. We know that hydrogen and helium do in fact make up about 99 per cent of the matter of the universe. This leaves us with the problem of building the heavier elements. I hold to the opinion that some of them were built by capture of neutrons. However, since the absence of any stable nucleus of atomic weight 5 makes it improbable that the heavier elements could have been produced in the first half hour in the abundances now observed, I would agree that the lion's share of the heavy elements may well have been formed later in the hot interiors of stars.

All the theories—of the origin, age, extent, composition and nature of the universe—are becoming more and more subject to test by new instruments and new techniques, which are described in later chapters. In the chapter on the red-shift investigations, Allan Sandage reports a tentative finding that the expansion of the universe may be slowing down. If this is confirmed, it may indicate that we live in a pulsating universe. But we must not forget that the estimate of distances of the galaxies is still founded on the debatable assumption that the brightness of galaxies does not change with time. If galaxies actually diminish in brightness as they age, the calculations cannot be depended upon. Thus the question whether evolution is or is not taking place in the galaxies is of crucial importance at the present stage of our outlook on the universe.

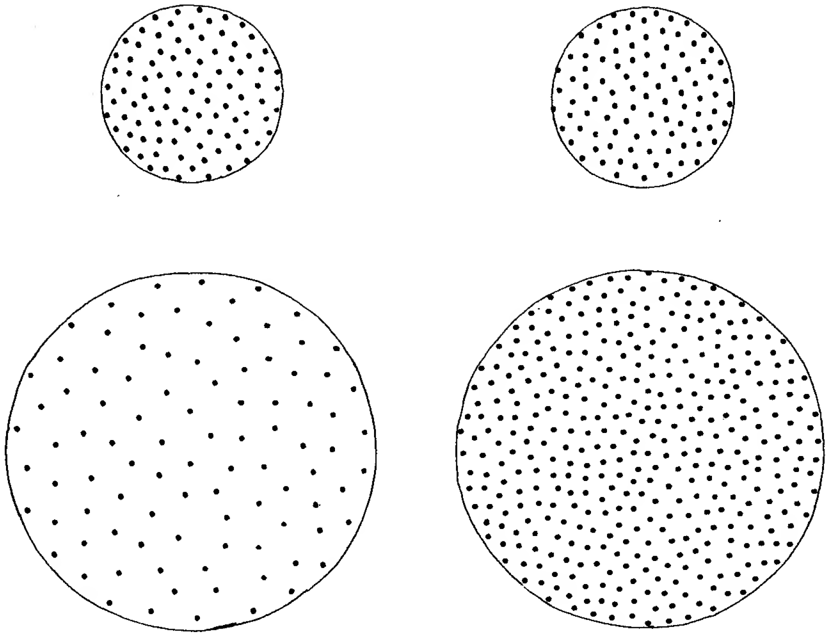
THE STEADY-STATE UNIVERSE

by Fred Hoyle

THE THEORY of a steady-state universe leads to many startling conclusions: that the universe had no beginning and will have no end, that space as well as time is infinite, that matter is continually being created throughout space—to mention a few. Human nature being what it is, there has been a tendency to become involved in emotional attitudes toward these concepts, instead of confining the discussion to purely scientific criteria. If the writer, along with critics, has transgressed in this respect, he promises to give some redress in this article.

The steady-state theory holds that the large-scale features of the universe do not change with time. Only the galaxies and clusters of galaxies change; if we "smear out" their material uniformly through space and consider the general properties of the cosmos, it is unchanging. The expansion of the universe is a basic feature of the theory. The question arises: If the galaxies are moving apart from each other, why does space not become more and more empty? The answer of the theory is that new galaxies and clusters of galaxies are constantly being formed, their rate of formation just compensating for the separating effect of the expansion. So a stable situation is preserved.

Before we go on to consider the reasoning, predictions and tests of steady-state cosmology, the writer should point out that his own approach to the theory, and also that of William H. McCrea of the University of London, differs rather markedly from the approach of Hermann Bondi and Thomas Gold. The writer's approach is a mathematical one developed in the framework of the theory of relativity. Bondi's and Gold's is founded on an intuitive but powerful physical principle. To understand



Evolutionary and steady-state views are compared in these diagrams. At left is a schematic view of an evolutionary universe. At the top is a sample of the universe, with the galaxies represented by dots. At the bottom is a picture of the same galaxies after the passage of time. The galaxies have merely receded from one another. At right is the same kind of view of a steady-state universe. At the top is a sample of the universe. At the bottom is a picture of the volume occupied by the same galaxies after the passage of time. In that time, however, new galaxies have been created, maintaining the density of the galaxies as before.

their outlook we must look into the nature of this postulate, which is called the "cosmological principle."

Cosmology differs from all other branches of physical science in a very important respect. Whereas other physical scientists deal always with isolated systems, whose "boundary conditions" can be defined, a cosmologist has to deal with a nonisolated system. To cope with this unhappy situation he is forced to adopt a "symmetry" postulate, which says that, local fluctuations apart, the universe will look the same from wherever one views it. That is to say, it assumes that observers attached to different galaxies anywhere in the cosmos would all obtain exactly the same large-scale picture of the universe. But if the universe changes with time, this implies that the different observers compare their pictures at the same time, which of course requires us to have a definition of what we mean by "at the same time." In order to make a definition of simultaneity possible, the mathematician Hermann Weyl advanced the additional postulate that the motions of the galaxies follow a regular type of pattern, whose exact nature need not be described here.

Instead of this additional postulate Bondi and Gold proposed a single all-embracing "cosmological principle": namely, that the large-scale features of the universe are the same not only from every point of view in space but also from every point of view in time. This symmetry hypothesis leads immediately to the conclusion that the universe is in a steady state. It is then immaterial whether the observers compare their pictures "at the same time" or not.

The outlook of Bondi and Gold has a compelling simplicity. Moreover, symmetry postulates have repeatedly demonstrated their power in theories of physics during the present century (e.g., the positive and negative particles of nuclear physics). But to my own taste it is preferable to start with a mathematical definition of the continuous creation of matter within the framework of the relativity theory and then to derive the steady-state solution as a consequence of field equations.

At first sight the creation of matter may seem a queer concept to be invading scientific thought. But as other chapters in this book make abundantly clear, the origin of matter must enter all cosmologies. Nowadays we are coming more and more to realize that hydrogen is the original material—the material out of which the other elements have been produced by nuclear reactions inside stars. This transmutation of hydrogen is going on all the time.

Why is there any hydrogen remaining in the universe? Why was it not all used up in the production of heavy elements eons ago? If we assumed that the hydrogen of the universe has existed for an infinite time, there would be two conceivable answers. We might suppose that the hydrogen has not had sufficient time to become transmuted into other elements because the stars were born only recently, that is, within the last five billion years or so. But it would follow from this that the hydrogen remained stable for eons of time and then suddenly five billion years ago began to condense into stars and galaxies. This seems less than plausible. The other possibility, assuming the hydrogen is infinitely old, is that we still find it on hand because the higher elements formed from it break down to hydrogen again. The chief objection to this idea is the difficulty of explaining how the energy necessary for the breakdown would be supplied. Decomposition of the heavier elements into hydrogen requires absorption of energy—the reverse of the release of energy that occurs when hydrogen nuclei combine. To provide an amount of energy adequate to account for a sufficiently large-scale reconversion of the heavier elements, nothing less than an implosion of the whole universe (as opposed to an explosion) apparently would suffice.

We are thus led to the conclusion that the hydrogen we observe is not infinitely old: it has originated within some finite time and has not yet been converted to heavier elements. Both the evolutionary and the steady-state theories of the universe agree on this point. But there the similarity between them ends. The evolutionary theory argues that all the hydrogen was created in an

explosive beginning some five and a half billion years ago (see the preceding chapter). The steady-state hypothesis holds that hydrogen has been created at a steady rate throughout infinite time and is still being created at the same rate today.

If hydrogen has been present for an infinity of time, and has steadily been converted to heavier elements in stars, why don't we see galaxies made of very old matter? Why do we see only comparatively young galaxies, composed almost entirely of hydrogen? The answer of the steady-state theory is that the expansion of the universe spreads galaxies apart as they age, and the old material is rapidly diluted, in terms of its mean density in the universe as a whole. Meanwhile new hydrogen, and new galaxies, are just as rapidly being created. According to the mathematics of the theory, the expansion of the universe and the creation of new material go on at rates such that the mean density of 200-billion-year-old material, for example, is less than that of recently formed material by a factor of 10^{43} (1 followed by 43 zeros). It must be emphasized that this figure is a mean averaged over the universe as a whole: it does not apply to individual galaxies or clusters of galaxies. Expansion takes place in space *between* galactic systems: the individual galaxies and clusters do not themselves expand. The very old material of the universe is concentrated in very old galaxies. By virtue of the universal expansion these are now extremely far apart. Possibly there are some moderately old galaxies within the range of our telescopes. If a method could be worked out to identify distant galaxies composed of comparatively old matter, it would provide a test of the steady-state theory.

Approaching the steady-state theory from the mathematical point of view, our first step evidently must be to construct a mathematical law representing the origin of matter. We wish to formulate this law within the logical framework of the theory of relativity: like the evolutionary theory, steady-state cosmology makes use of the powerful equations devised by Albert Einstein

to describe the four-dimensional space-time continuum. We can indicate briefly here some of the main principles involved, though the equations themselves are too complex to examine in detail.

The theory of relativity begins by generalizing the ordinary laws of motion in three-dimensional space to describe the properties and the non-Euclidean geometry of the four-dimensional space-time field. These laws can be set down in four equations: one equation for the law of conservation of energy and three for the conservation of momentum. Our problem is to frame the law of origin of matter in such a way that it can be introduced into these four conservation equations.

As a first step we must define energy and momentum, for the theory of relativity does not itself define them. It is most reasonable to choose definitions which will yield equations as closely analogous as possible to the ordinary equations describing the laws of conservation in our familiar (Euclidean) world. The evolutionary cosmologists seem at first sight to have done this, but it turns out that their conservation equations do not contain any generalized analogue of certain terms, known as "fluid stresses," which play a part in the ordinary equations. Now when we define energy and momentum in a way that yields such a generalization, the outcome of the equations is a steady-state universe, not an evolutionary one.

The equations, so generalized, imply a "feedback" relation between the expansion of the universe and the origin of matter. If the expansion rises above a certain critical rate (related to the rate of origin of matter), the feedback slows the expansion. If the universe's expansion slows down to less than the critical rate, the feedback speeds it up. Thus the interaction between the expansion and the creation of matter maintains a steady state in which the mean density of matter in the universe remains constant.

To many people the notion of continuous creation of new matter in space seems an outright violation of the conservation of energy. But this indicates a confusion between a closed system

and the very different situation in an open system. The theory of relativity says that in an open, infinitely expanding universe, local concentrations of energy are related to the energy of expansion of the whole universe. The energy of expansion can take a form leading to a continuous creation of local matter.

The same question that is asked about the creation of matter might be asked about the red shift of light from distant galaxies. The reddened light is weaker than when it started on its journey. Where does the lost radiant energy go to? It goes into a slight increase in the rate of expansion of the whole universe. The point is that for a total reckoning of the conservation of energy and matter in the cosmos we must take the expansion of the universe into account. We cannot balance the energy books strictly and completely within the confines of any locality, because no locality in the universe is entirely closed.

Before we drop this issue it is perhaps worth noting that we can consider the conservation question in purely operational rather than theoretical terms and come out with the same result. Suppose observers on the earth measured the energy content of a given portion of the universe, say that within the reach of the 200-inch telescope, and suppose this was done on several occasions at widely separated times. If the conservation of energy is to hold, the measured energy content must remain unchanged from one occasion to another. This would be true in a steady-state universe, but not in an evolving one. Furthermore, in a steady-state universe conservation in this operational sense holds good for an observer in any galaxy.

The two features of the steady-state theory that seem to cause the greatest general surprise are (1) that the theory possesses a clear cut mathematical basis, and (2) that the theory is highly susceptible to test by observation. How can it be tested? Obviously we cannot test it in the laboratory—unless we were to find some way to speed up the creation of matter artificially—for the rate of creation, according to the theory, is negligible in terrestrial

terms: in the space of the average physics laboratory one new hydrogen atom would materialize in about 1,000 years. But on the cosmological scale there are many possible tests.

First, at the farthest range of our telescopes we are seeing galactic systems as they were a billion or more years ago. Hence information can be obtained about how things used to be in the past, and this information can be compared with the cosmic scene close by us in space and time. Since the steady-state theory requires that there be no difference in large-scale properties between the past and the present, the theory is clearly exposed to check by this comparison. Large-scale properties can be estimated from many different clues: the density of the populations of galaxies, the magnitude and color of their light, the radio emissions signaling collisions and other significant events, the relation between the red shift and distance of galaxies, and so forth.

Second, there are tests which can be made without looking so far away from home. We can think of the formation of new galaxies as equivalent to birth in the biological sense, and of their separation by expansion as equivalent to death. In terms of this analogy a new generation of galaxies is born, not every 30 years as in the case of human beings, but every few billion years. Now, just as an animal population becomes extinct if it fails to reproduce its numbers from generation to generation, so large-scale properties of the systems of galaxies fail to survive unless they reproduce themselves in the same sense. If the universe is infinitely old, as the steady-state theory says, we should expect to see surviving only properties which have stabilized themselves so that they are reproduced at precisely the same level from generation to generation. In other words, according to the steady-state theory the galaxies are not a product of random fluctuations and condensations, as in the evolutionary theory, but represent a very strictly controlled system obeying a kind of cosmic ecology, with the origin of matter playing a critical role. This crucial difference between the two theories can form the basis of stringent tests. The tests can be applied to such properties as the density of galaxies in space and the distribution of sizes in masses of galaxies. That is, we can check whether the distribution follows a regular frequency curve or shows no regular pattern.

During the past five years it has twice been claimed that observations disproved the steady-state theory, but it now appears that in both cases the observations are open to serious doubt. The United States astronomers Joel Stebbins and A. E. Whitford thought that certain distant galaxies showed more reddening than could be attributed to the usual red shift, and this was construed to support the evolutionary theory. But Whitford later found that certain data they had made use of were incorrect. Recently Martin Ryle in England reported a count of radio sources which indicated that the density of galaxies in space increases with distance from us—again an apparent support for the evolutionary hypothesis. However, Ryle's findings have been questioned by the radio astronomer B. Y. Mills in Australia.

In my view the most serious potential contradiction of the steady-state theory lies in the recent red-shift studies by the astronomers in California, which are reported in this book by Allan Sandage (see the following chapter). As Sandage points out, however, the findings so far are highly uncertain.

George Gamow has offered against the steady-state theory the objection that elliptical galaxies (which are thought to consist only of old stars) apparently do not show the age variations that the theory predicts. In defense of the theory it can be said that the measurements cited (studies of the color of the galaxies, in two colors) are not a very sensitive index of the galaxies' ages. In the color test a galaxy six billion years old should look much like one three billion years old. More sensitive measurements are required.

The steady-state theory gains support, on the other hand, from recent studies indicating that the elements beyond hydrogen are formed in stars. These studies, reported in the chapter by

William A. Fowler (page 17), make it appear more likely that the elements are constantly being "cooked" in the stars, as the steady-state cosmology suggests, than that they were created in a primeval explosion, as Gamow has urged.

Radio astronomy offers the exciting possibility of something close to a direct test of the creation of matter in space. The total amount of matter in the galaxies we can observe is estimated to come to about 10^{-30} of a gram per cubic centimeter if it were spread evenly all through space. The steady-state theory predicts that the average density of matter should be 10 or more times greater than this. The difference, according to the theory, is accounted for by hydrogen spread through intergalactic space. Up to now it has not been possible to detect intergalactic matter. But in the next few years new radio telescopes, tuned to the one-note "song of hydrogen," may be able to test whether such quantities of hvdrogen do or do not exist in space.

PART 5 OBSERVATIONAL TEST

I. THE RED SHIFT *by Allan R. Sandage*

The author is assistant astronomer at the Mount Wilson and Palomar Observatories. He acquired his doctorate in astronomy at the California Institute of Technology in 1953, having meanwhile joined the staff of Mount Wilson and Palomar. Few astronomers have set out on their professional careers so early in life. "As I recall, it was *Buck Rogers in the 25th Century* that steered me into astronomy at the unknowing age of 10. This unfortunate interest (from our neighbors' point of view) took the form of dragging a telescope, tables and other observing paraphernalia into our back yard at three in the morning to look at meteor showers and the like. From 1951 to 1953 I was an assistant to the late Edwin P. Hubble. His principal interest was observational cosmology, and a bit of his enthusiasm rubbed off on all who knew him. The associations with Hubble, Milton L. Humason and Walter Baade have been the high points in my scientific career. At present my chief interest is in a slightly different field: that of the observational approach to stellar evolution."

II. THE DISTRIBUTION OF GALAXIES *by Jerzy Neyman and Elizabeth L. Scott*

A statistician temporarily turned astronomer and an astronomer turned statistician collaborated on the story told in this chapter. Jerzy Neyman is professor and director of the Statistical Laboratory at the University of California. Elizabeth Scott is assistant professor. Neyman was born in Bendery, Rumania; he was trained in mathematics and statistics at the universities of Kharkov, Warsaw, London and Paris. He came to the University of California in 1938, a year before Miss Scott, a native of Oklahoma, got her bachelor's degree there.

How do statisticians get immersed in astronomy? Neyman explains: "For me personally cosmological problems are a relatively novel field. Up to 1950 I certainly did not think about them at all. By then C. D. Shane of the Lick Observatory had already completed a substantial amount of his survey of the sky and was struck by certain regularities in the general chaos of the distribution of images of galaxies on his plates. He paid us frequent visits in the Statistical Laboratory and discussed with us a number of problems in which he appeared to be intensely interested. Some of these problems were familiar to Miss Scott because she was originally an astronomer. Shane's enthusiasm was pretty soon matched by that of Miss Scott and, in due course, I also became fascinated."

III. RADIO GALAXIES *by Martin Ryle*

The author is a research fellow of Trinity College in the University of Cambridge. After graduating from the University of Oxford on the eve of World War II, he plunged into radar research and development for the Royal Air Force. In 1945 he received an appointment to the Cavendish Laboratory at Cambridge and there, with D. D. Vonberg, initiated radio observations of the sun. "You will see that I came to radio astronomy from a training in physics, via a period of intensive work in radio and radar engineering. Part of the latter was concerned with problems of aerial design, which experience has of course stood me in good stead in radio-telescope design. Although my direct approach has been by way of radio engineering, astronomy was an early interest." Since 1952 Ryle has been a Fellow of the Royal Society.

THE RED SHIFT

by Allan R. Sandage

In the nature of things it is a delicate undertaking to try to discern the general structure and features of a universe which stretches out farther than we can see. For more than a quarter of a century both the theoreticians and the observers of the cosmos have been making exciting discoveries, but the points of contact between the discoveries have been few. The predictions of the theorists, deduced from the most general laws of physics, are not easy to test against the real world—or rather, the small portion of the real world that we can observe. There is, however, one solid meeting ground between the theories and the observations, and that is the apparent expansion of the universe. Other aspects of the universe may be interpreted in different ways to fit different theories, but concerning the expansion the rival theories make unambiguous predictions on which they will stand or fall. There is now hope that red-shift measurements of the universe's expansion with the 200-inch telescope on Palomar Mountain will soon make it possible to decide, among other things, whether we live in an evolving or a steady-state universe.

Let us begin by considering just what issue the measurements seek to decide, as between the theories expounded in the two preceding chapters by George Gamow and Fred Hoyle. The steady-state theory says that the universe has been expanding at a constant rate throughout an infinity of time. The evolutionary theory, in contrast, implies that the expansion of the universe is steadily slowing down. If the universe began with an explosion from a superdense state, its rate of expansion was greatest at the beginning and has been slowing ever since because of the oppos-

ing gravitational attraction of its matter, which acts as a brake on the expansion—much as an anchored elastic string attached to a golf ball would act as a brake on the flight of the ball.

Now in principle we can decide whether the rate of expansion has changed or not simply by measuring the speed of expansion at different times in the universe's history. And the 200-inch telescope permits us to do this. It covers a range of about two billion years in time. We see the nearest galaxies as they were only a few million years ago, while the light from the most distant galaxies takes so long to reach us that we see them at a stage in the universe's history going back to one or two billion years ago. If the explosion theory is correct, the universe should have been expanding at an appreciably faster rate then than it is now. Since the light we are receiving from the distant galaxies is a flashback to that earlier time, its red shift should show them receding from us faster than if the rate of expansion had remained constant.

The red shift is so basic a tool for testing our notions about the universe that it is worth while to review how it was discovered and how it is used.

An astronomer cannot perform experiments on the objects of his study, or even examine them at first hand. His information rides on beams of light from outer space. By sufficiently ingenious instruments and equally ingenious interpretation (we hope), he may translate this light into information about the temperatures, sizes, structures and motions of the celestial bodies. It was in 1888 that a German astronomer, H. C. Vogel, first demonstrated that the spectra of stars could give information about motions which could not otherwise be detected. He discovered the Doppler effect in starlight.

The Doppler effect, as every physics student knows, is a change in wave length observable when the source of radiation (sound, light, etc.) is in motion. If it is moving toward the observer, the wave length is shortened; if away, the waves are lengthened. In the case of a star moving away from us, the

whole spectrum of its light is shifted toward the red, or long-wave, end (see Plate 17).

This spectrum, made by means of a prism or diffraction grating which spreads the light out into a band of its component colors, is usually not continuous. Certain wave lengths of the light are absorbed by atoms in the star's atmosphere. For example, most stars show strong absorption, by calcium atoms, at the wave lengths of 3933.664 and 3968.470 angstrom units. (An angstrom unit is a hundred-millionth of a centimeter.) The absorption is signaled by dark lines in the spectrum, known in this case as the K and H lines of calcium. Now, if a star is moving away from us, these lines will be displaced toward the red end of the spectrum. In the spectrum of the star known as Delta Leporis, for instance, the K line of calcium is displaced 1.298 angstroms toward the red. Assuming the displacement is due to the Doppler effect, it is a simple matter to calculate the velocity of the star's receding motion. Dividing the amount of the displacement by the normal wave length at rest, and multiplying by the speed of light (300,000 kilometers per second) we get the speed of the star—in this case 99 kilometers per second. The calculation on the basis of displacement of the H line gives the same figure.

Equipped with this powerful tool, many of the large observatories in the world spent a major part of their time during the early part of this century measuring the velocities of receding and approaching stars in our galaxy. At first it was a work of pure curiosity, no one suspecting that it might have any bearing on cosmological theories. But in the 1920s V. M. Slipher of the Lowell Observatory made a discovery which was to lead to a completely new picture of the universe. His measurements of red shifts of a number of "nebulae" then thought to lie in our galaxy showed that they were all receding from us at phenomenal speeds—up to 1,800 kilometers per second. Edwin P. Hubble at Mount Wilson soon established that the "nebulae" were systems of stars, and he went on to measure their distances. The method he used was the one developed by Harlow Shapley, employing Cepheid

variable stars as the yardstick. Shapley had found a way to measure the intrinsic brightness of these stars, and therefore their distance could be estimated from their apparent brightness by means of the rule that the intensity of light falls off as the square of the distance. Hubble observed that the galaxies nearest our own system, including the Great Nebula in Andromeda, contained Cepheid variables, and when he computed their distances he came out with the then astounding figure of about one million light-years! He next tackled the problem of finding the distances of Slipher's nebulae. Since variable stars could not be detected in them, he used their brightest stars as distance indicators instead. He found that these nebulae were at distances ranging up to 20 million light-years from us, and what was more remarkable, their velocities increased in strict proportion to their distances!

Hubble made the daring conjecture that the universe as a whole was expanding. He predicted that more remote galaxies would show larger red shifts, still in proportion to their distance. To test Hubble's speculation, Milton L. Humason began a long-range program of spectral analysis of more distant galaxies with the 100-inch telescope on Mount Wilson. In these faint galaxies it was no longer possible to distinguish even the bright stars, and so the relative brightness of the galaxy as a whole had to be taken as the measure of distance. That is, a galaxy one fourth as bright as another was assumed to be twice as far away. Hubble reasoned that while individual galaxies might deviate from this rule, statistically the population of galaxies as a whole would follow it. The principle is still the basis of distance determinations today.

Humason laboriously photographed spectra of galaxies, and Hubble measured their apparent brightness, from 1928 to 1936, when they reached the limit of the 100-inch telescope. The history of the red-shift program in those years is a story of extreme skill and patience at the telescope and of steady improvement in instrumentation. It was a long and difficult task to photograph spectra then; the prisms used required long exposures, and it took

10 nights or more to obtain a spectrum which with modern equipment can be recorded in less than an hour today. The improvement in equipment includes not only the 200-inch telescope but also diffraction gratings, faster cameras and a vast improvement in the sensitivity of photographic plates, thanks to the Eastman Kodak Company. Astronomers the world over, and cosmology, owe a large debt to the Eastman research laboratories.

Humason's first really big red shift came early in 1928, when he got a spectrum of a galaxy called NGC 7619. Hubble had predicted that its velocity should be slightly less than 4,000 kilometers per second: Humason found it to be 3,800. By 1936, at the limit of the 100-inch telescope's reach, they had arrived at a cluster of galaxies, called Ursa Major No. 2, which showed a velocity of 40,000 kilometers per second. All the way out to that range of more than half a billion light-years the velocity of galaxies increased in direct proportion to the distance. In a sense this was disappointing, because the various cosmological theories predicted that some change in this relation should begin to appear when the observations had been pushed far enough. Further exploration into the distances of space had to await the completion of the 200-inch Hale telescope on Palomar Mountain.

In 1951 the red-shift program was resumed, with a new spectrograph of great speed and versatility placed in the big telescope's prime focus cage, where the observer rides with his instruments. The spectrograph has to be of very compact design to fit into the cramped space of the cage. The photographic plate itself, mounted in the middle of a complex optical arrangement, is only 15 millimeters (about half an inch) on a side. The cutting and handling of such small pieces of glass in complete darkness (to avoid exposure of the plate) is a tricky business. The spectrum recorded on the plate is a tiny strip only a fifth of an inch long, but it is long enough to measure red shifts to an accuracy of better than one half of one per cent.

The most distant photographable galaxies are so faint that they are not visible to the eye through the telescope: they can be

recorded only by extended exposure of the plate. The observer guiding the telescope must position the slit of the spectrograph by reference to guide stars within the same field as the distant object. Another great difficulty in recording the red shift of extremely distant galaxies arises from the magnitude of the shift. The displacement of the calcium dark lines toward the red is so large that the lines move clean off the sensitive range of blue photographic plates, which astronomers like to use because of their speed. So slow panchromatic plates must be used, and Humason has been forced to return to exposure times as long as 30 hours or more.

The other part of the program—measuring the distances of the galaxies—also has been helped by improvements in technique. For measurement of their brightness the Mount Wilson telescopes employ photomultiplier tubes, which amplify the light energy by electronic means. Such equipment was not available for the 200-inch telescope when the present program began. Instead the intensity of the light from very faint galaxies was measured by a tricky method which compares it with that of stars of known magnitude. No direct comparison can be made, of course, between the picture of a star and that of a galaxy or cluster of galaxies, because the star is a point source of light while a galactic system is a spread-out image. To make the images comparable, a region of the sky is photographed with a "jiggle" camera which moves the plate around so that the images of stars and of galaxies are smeared out in squares. They can then be compared as to brightness—just as one may use color cards to find a match to the color of a room.

Humason has now measured red shifts of remote clusters of galaxies with recession velocities up to 60,000 kilometers per second. What do they show? Is the velocity still increasing in strict proportion to the distance?

The information about 18 of the faintest measured clusters is given in the chart on page 97. Their velocities are plotted against

their apparent brightness, or estimated distances. If velocity increases in direct proportion to the distance, the observed velocity-distance relation should be "linear" (i.e., follow a straight line). But as the chart shows, the very faintest clusters have begun to depart from that line. These clusters, about a billion light-years away, are moving *faster* (by about 10,000 kilometers per second) than in direct proportion to their apparent distance. In other words, the data would be interpreted to mean that a billion years ago the universe was expanding faster than it is now. If the measurements and the interpretation are correct, this suggests that we live in an evolving rather than in a steady-state universe.

The observed change in the curve buys us much more information. To begin with, it tells us something about the mean density of matter in the universe. The rate at which the expansion of the universe is slowing down (if it is) depends on the mean density of its matter: the higher the density, the greater the braking effect. The amount of departure from linearity indicated by the measurements thus far calls for a mean density of about 3×10^{-28} grams of matter per cubic centimeter (about one hydrogen atom per five quarts of space). Now, this amounts to about 300 times the total mass of the matter estimated to be contained in galaxies: that figure comes out to a mean density of only 10^{-30} grams per cubic centimeter. If our present tentative value for the slowdown of the expansion should be confirmed, we would have to conclude that either the current estimates of the masses of the galaxies are wrong or that there is a great deal of matter, so far undetected, in intergalactic space. Matter in the form of neutral hydrogen (i.e., normal hydrogen atoms consisting of a proton and an electron) might be present in space and still have escaped detection until now because it is not luminous. The giant radio telescopes now under construction or on the drawing boards perhaps will detect the hydrogen, if it exists in the postulated quantities.

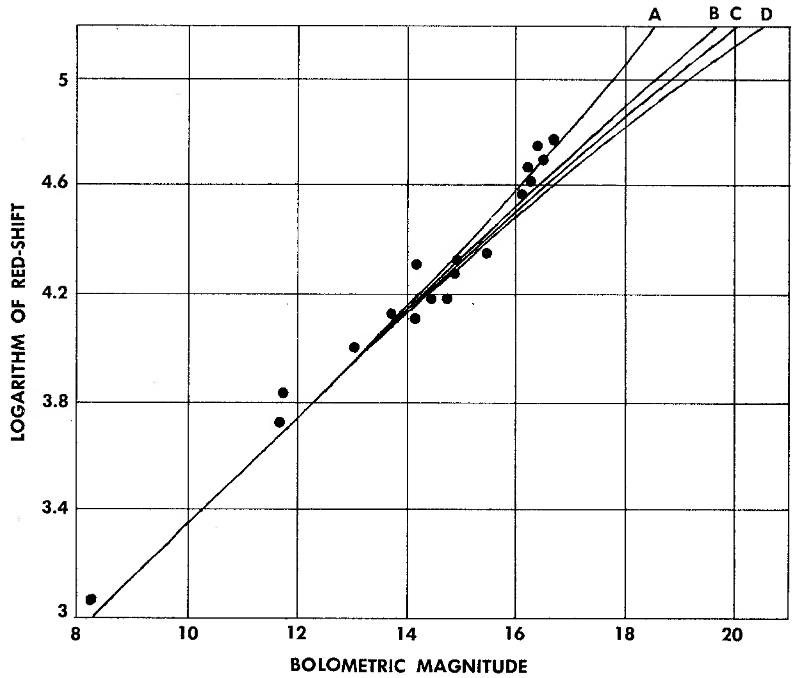
Once we know the rate at which expansion of the universe is

slowing down, it becomes possible to determine not only the mean density of matter but also the geometry of space—that is, its curvature. As George Gamow points out in the chapter on evolutionary cosmology (page 59), models of the evolving universe take three forms: the Euclidean case, in which space is flat, open and infinite; a curved universe which is closed and finite, like the surface of a sphere; and a curved universe which is open and infinite, like the surface of a saddle. In the accompanying velocity-distance chart (page 97) curves to the left of C represent evolving models, and curve D represents the steady-state model. If the curve of the velocity-distance relation lies between C and B, the universe is open and infinite. Line B is the Euclidean case of flat space. If the curve is left of B, the universe is closed and finite, the radius of its curvature decreasing as we move farther to the left.

According to our present observations, the actual relation follows a curve left of B (curve A on the chart). Although our data are still crude and inconclusive, they do suggest that the steady-state model does not fit the real world, and that we live in a closed, evolving universe.

Humason has gone beyond 60,000 kilometers per second and attempted to photograph and measure the red shifts of two faint clusters whose predicted velocity is more than 100,000 kilometers per second. So far these efforts have not yielded reliable results, but he is continuing them. Also, William A. Baum of the Mount Wilson and Palomar Observatories, using photoelectric equipment, has detected an even more distant cluster with a velocity of 120,000 kilometers per second. Baum's measurement gives a smaller degree of curvature than the photographic data have sugested, but his point is still slightly to the left of curve B, and thus also indicates a closed, finite universe. The red-shift program is continuing toward the goal of definitely determining the geometry of space.

If the expansion of the universe is decelerating at the rate the photographic data suggest, the expansion will eventually stop



Eighteen faintest clusters of galaxies yet measured are plotted for their red shift (or speed of recession) and apparent magnitude (or distance). Line *C* represents a universe expanding forever at the same rate. Line *D* is a steady-state universe. If the measured line falls to the left of *C*, the expansion must slow down. If it falls between *C* and *B*, the universe is open and infinite. If it falls to the left of *B*, the universe is closed and finite. If it falls on *B*, it is Euclidean and infinite. *A* is the trend suggested by the six faintest clusters.

and contraction will begin. If it returns to a superdense state and explodes again, then in the next cycle of oscillation, some 15 billion years hence, we may all find ourselves again pursuing our present tasks.

Although no final answers have yet emerged, big steps have been taken since 1928 toward the solution to the cosmological problem, and there is hope that it may now be within our grasp. The situation has nowhere been better expressed than in Hubble's last paper:

"For I can end as I began. From our home on the earth we look out into the distances and strive to imagine the sort of world into which we are born. Today we have reached far out into space. Our immediate neighborhood we know rather intimately. But with increasing distance our knowledge fades . . . until at the last dim horizon we search among ghostly errors of observations for landmarks that are scarcely more substantial. The search will continue. The urge is older than history. It is not satisfied and it will not be suppressed."

THE DISTRIBUTION OF GALAXIES

by Jerzy Neyman and Elizabeth L. Scott

IN THE EFFORT to obtain a large-scale view of the universe, most cosmological studies are forced to reduce its observed make-up and behavior to averages—one atom to so many quarts of space, "average" galaxies set out uniformly in space like orange trees in a grove. This smoothing-out procedure is unavoidable in any attempt to describe the universe in terms of cause-and-effect relationships. The motion of a single planet may be predicted accurately by a few formulas, but to consider a whole universe of stars and galaxies in this way is completely out of the question. There are just too many bodies requiring separate equations, and the equations become far too complicated to handle. The formulas of the cosmologies are therefore applied to an averaged picture to see what general conclusions can be drawn. This approach, however, has obvious limitations. The conclusions can agree with the real universe only in terms of averages, and they leave unexplained many of the details of its structure and behavior.

There is another way of attacking the problem. We may give up the cause-and-effect approach and consider the universe as an outcome of a chance mechanism, subject to the laws of probability. A roulette wheel is a chance mechanism. The significant feature of such mechanisms is that they produce striking regularities "in the large" combined with a tremendous range of irregularities "in the small." Thus an appropriate mechanism might reproduce the large regularities and the pattern of local irregularities of the universe as we know it. In other words, we would seek to re-create the universe we see by the operation of some chance mechanism applied repeatedly inch by inch over all

space and hour by hour over all time. Such a mechanism could not attempt to predict exactly what would happen at a given moment of time in a given region of space. But it would try to predict how frequently a given configuration of stars, galaxies or other systems will be found in different regions in space.

We have been working along these lines with our colleagues at the University of California and the Lick Observatory. C. D. Shane, astronomer at the Observatory, prompted our study. He was engaged in a survey of the distribution of galaxies in photographs of the sky, and he became curious to know whether their lumpy distribution (rather like handfuls of seed scattered in a field) might be described by some statistical law. Our efforts to answer his question developed into a long-range study. The results so far obtained include a plausible chance mechanism governing the distribution of galaxies in space and a novel method of testing certain cosmological models. The method is independent of the red shift; indeed it offers an opportunity for a separate check on whether the universe is truly expanding. It also seems capable of deciding between the evolutionary and steady-state theories.

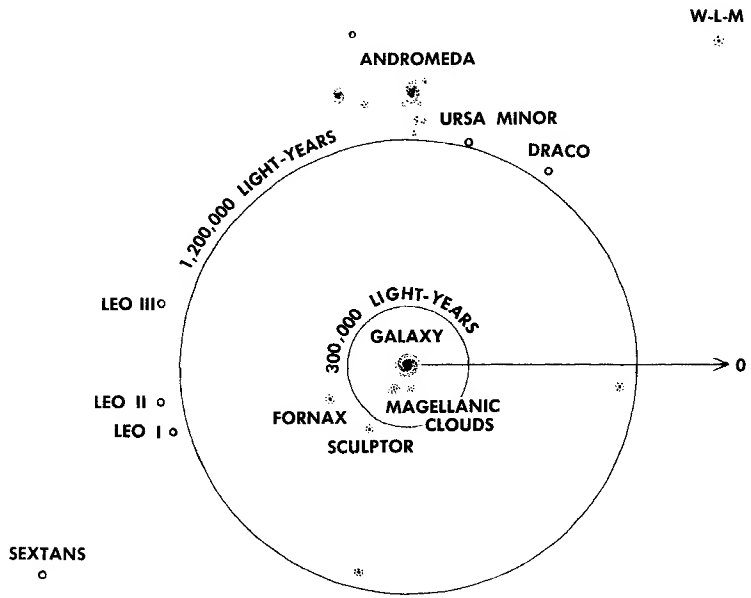
For more than a century astronomers have been making systematic surveys of the distribution of galaxies in the sky: several are in progress just now. No general pattern of any kind has emerged. The galaxies are found in small groups, in large clusters, in clusters of clusters, and alone in wide space. A rough picture of the nature of their distribution is given in the two maps which follow on pages 102 and 103. The first is a "close-up" view of the arrangement of galaxies in our own neighborhood, covering a span of about one million light-years. The second is a wider picture on a different scale which reduces all the galaxies in our local group to a dot in the center. It embraces a space of some 400 million light-years. On this scale no individual galaxies are shown but only clusters of them (which within the range of 150 million light-years can be distinguished into small "groups" and larger "clusters").

The maps illustrate the patchiness of the distribution of galaxies in space. It is this sort of picture that the chance mechanism procedure we have mentioned seeks to reproduce.

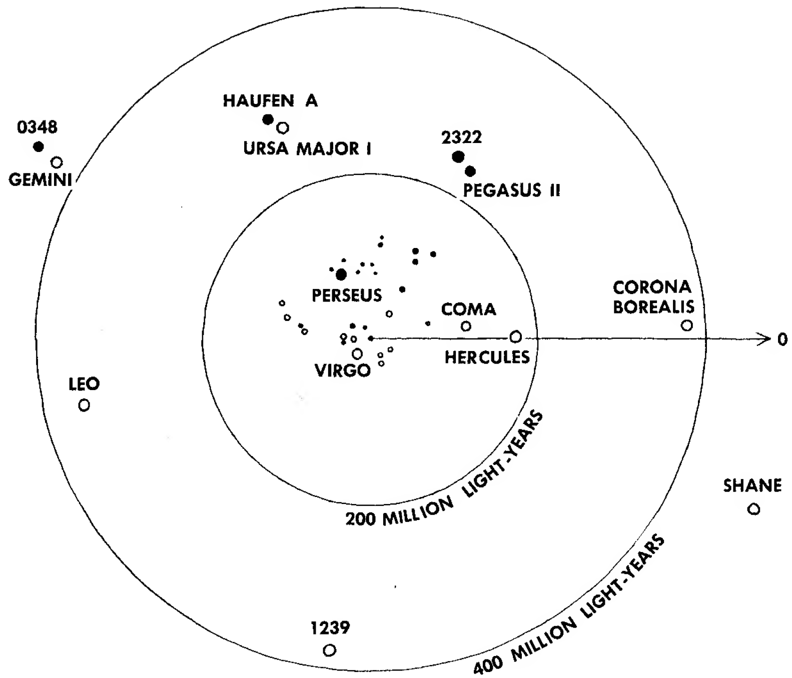
To explain our approach it may be helpful to consider as a rough analogy the problem of a life insurance actuary constructing a mortality table. His "universe" is a group of, say, one million policyholders, and his problem is to predict the state of this universe at various times in the future. Now, the fate over the next 12 months of a single human being, or of a family, is unpredictable, but if the actuary considers the million policyholders as a group, regularities "in the large" emerge: the proportion of deaths is greater among the 50-year-olds than among the 20-year-olds, etc. The whole process of survival and death occurs as if it were governed by a chance mechanism which can be represented as follows. Each morning before breakfast every single one of us approaches an urn filled with white and black balls. We draw a ball. If it is white, we survive the day. If it is black, we die. The proportion of black balls in the urn is not the same for each day, but grows as we become older in accordance with the so-called Gompertz-Makeham law. Still there are always some white balls present and some of us continue to draw them day after day for many years.

Naturally this chance mechanism is a hypothetical one. However, the important point is that by using such a hypothetical chance mechanism actuaries can predict the frequencies of the various combinations of disasters that may occur in a family and provide us with insurance. In other words, all happens as if we actually did draw balls before breakfast.

It is plausible that a similar chance mechanism, functioning repeatedly over small units of space and time, would produce the observed large-scale regularities in the distribution of matter in the universe and the correct frequencies of the various local irregularities. In parallel with the division of the life span of an individual into days, we visualized space as divided into an infinity of elements of volume (small cubes, all of the same size).



Local group of galaxies is roughly projected in the plane of the page. Our galaxy is in the center; the arrow indicates zero degrees on the Milky Way. The form of the galaxies is schematically shown, but their size is exaggerated. The objects to the north of the Milky Way are shown as open circles; those to the south of the Milky Way, as solid circles. The galaxies are named according to various conventions.



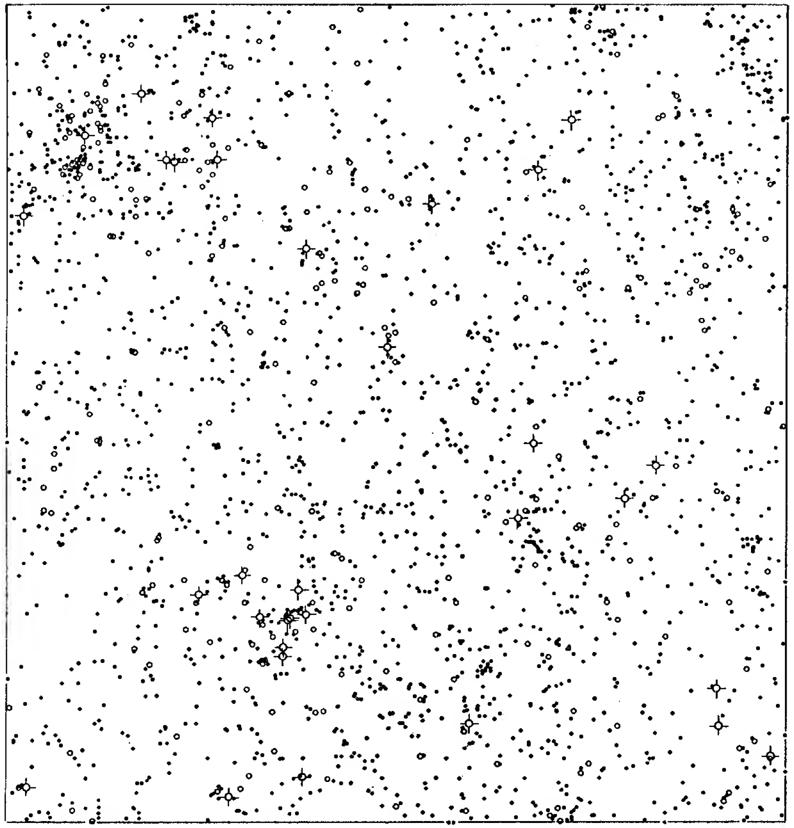
Associations of galaxies far beyond the local group are similarly mapped. The smaller dots represent groups of 50 galaxies or less; the larger dots, clusters of more than 50 galaxies. On this scale the entire local group is in the small central dot. The open circles are again those to the north of the Milky Way; the closed circles, those to the south of the Milky Way.

Just as each day is considered as a potential date of death, each elementary volume in space was treated as a potential location of a galaxy.

The simplest attack is to try to populate the universe by a single mechanism applied uniformly throughout space, and this is the one with which we began. Suppose that the chance mechanism is a roulette wheel with a very large, and specified, number of slots. Approach the first cube and spin the wheel. If the ball stops in the slot labeled O, put a galaxy in the cube. Otherwise leave it empty. Repeat the procedure for all the cubes. (To keep the mathematics manageable we consider space to be Euclidean—not curved—and infinite.) Carried out through all space, this process would produce some kind of arrangement of galaxies. By mathematical calculations it is possible to determine what the distribution of galaxies on photographs of the sky would be like if such a mechanism functioned uniformly through space.

When the calculated scheme of distribution was compared with the actual distribution of galaxies recorded in Shane's photographs of the sky (opposite page), it became apparent that the simple mechanism postulated could not produce a distribution resembling the one we see. In the real universe there is a much more pronounced tendency for galaxies to be grouped in clusters.

This suggested a different approach to the game. Suppose we assume that galaxies always occur in clusters, varying in the number of members. In principle we can consider even an isolated galaxy as a cluster containing but one member. We can then use a more elaborate chance mechanism to represent this situation. Approach the first cube, as before, and spin the roulette wheel. If the ball hits O, place in it not a galaxy but a "cluster center." After cluster centers have been placed by this random process in the whole space, we go back to the cluster centers to decide how many galaxies each cluster should contain. For each center we spin a roulette wheel again—this time a different wheel with a much smaller number of slots. If O comes up on the first spin,



Plot of galaxies was prepared by C. D. Shane from a photographic plate exposed at the Lick Observatory. It covers a square area about five degrees on a side. The sizes of the various symbols are a rough index of the brightness of the galaxies which they represent. If clustering is present, it cannot be definitely recognized by simple inspection.

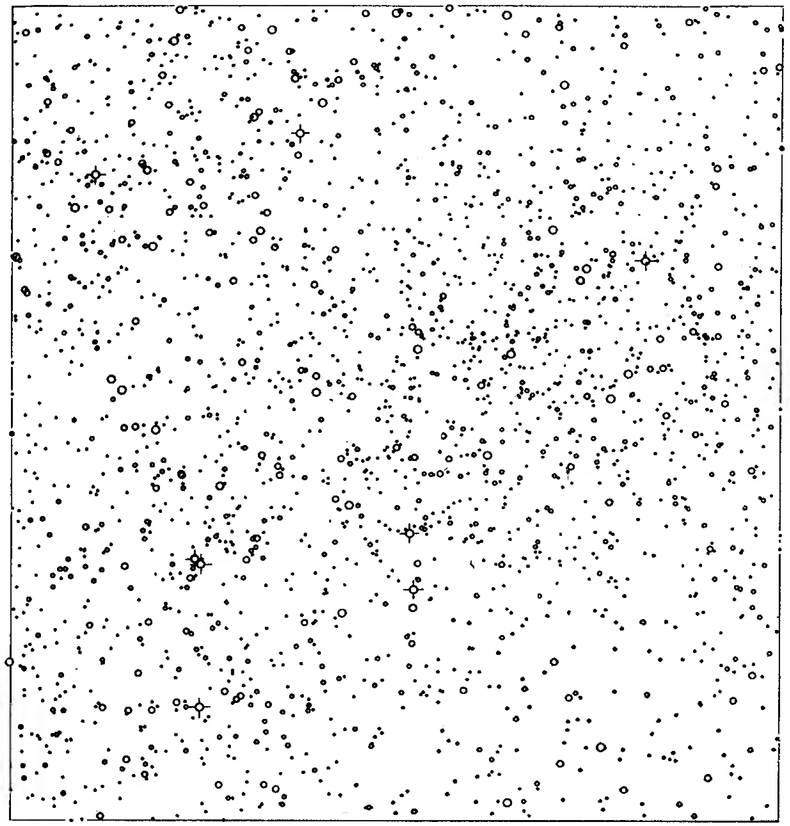
we place just one galaxy in the cluster. If it does not, we spin the wheel repeatedly until it comes up: the number of spins needed to hit O is taken as the number of galaxies in the cluster. Finally, we must arrange the galaxies in each cluster, deciding how far apart they shall be. For this we select a third appropriate roulette wheel.

It is obvious that by trying many different chance mechanisms and combinations of mechanisms (e.g., adding a fourth step to create clusters of clusters), we can produce a great variety of different distributions of galaxies. To determine the constants (i.e., the number of slots in each roulette wheel) that give the best fit with the observed universe is a tedious mathematical process. It involves hundreds of long and repetitive computations on the probability equations that represent the "game." These are best done on a large computer. On the observational side, the distribution of galaxies in photographs of the sky must be analyzed painstakingly to disclose their statistical anatomy for comparison with the chance-mechanism picture.

Some of this work has been finished and we have a tentative set of values for the constants which produces a distribution of galaxies of a kind similar to the one actually observed on photographic plates.

In the meantime a method of testing cosmological theories by studying the sky statistically has emerged. It rests upon the fact that the distribution we see is not a picture at one moment in time but represents distributions at widely different periods in the history of the universe. The more remote clusters are showing their distribution, or density, as it was hundreds of millions of years ago, because of the time it has taken their light to reach us.

The chance-mechanism process we have described for reproducing the universe assumes that the whole operation is carried out instantaneously. Thus any picture synthesized by this process is a picture of the universe at a single instant in time. If the universe is static (not expanding) or in a steady state, the density

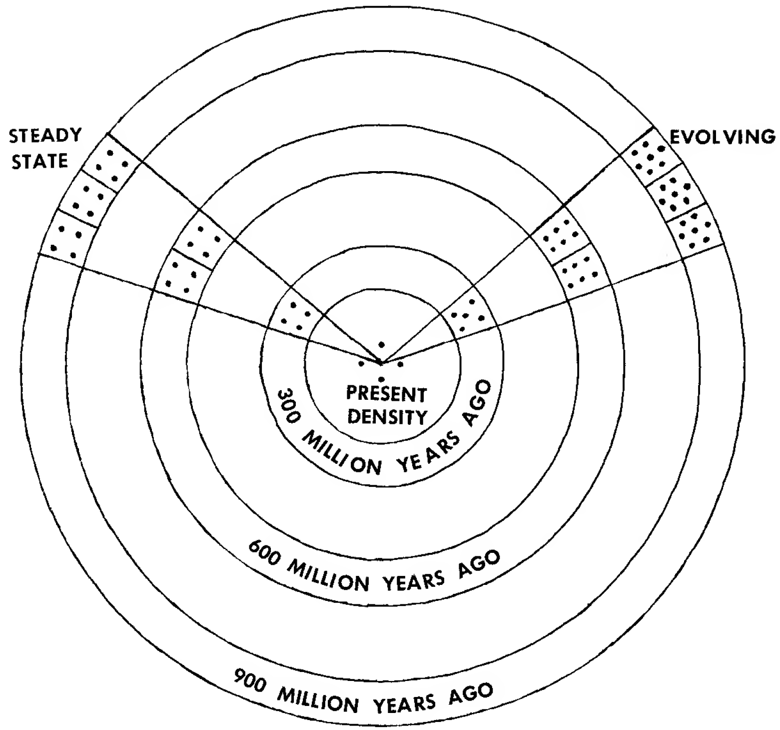


"Synthetic" plot of galaxies was made by actually applying the chance mechanism described in the text. The clustering of galaxies, which this mechanism guarantees, is also not apparent from a casual inspection. It is brought out only by statistical analysis.

of galaxies in space should be the same, of course, however far back we look in time. But if the universe is expanding and no new matter is being created, the clusters of galaxies are more spread out now than they were in the past. Therefore in the far reaches of space, where we see a flashback to the population of space as it existed hundreds of millions of years ago, the clusters of galaxies should be closer together than they are now. That is to say, a given volume of "old," distant space should contain more clusters than the same volume of space near us (see diagram opposite).This indicates a test of the theories: if the density of clusters in space increases with distance from us, the universe is expanding and evolving; if the density remains constant, the universe is static or in a steady state.

How can we compare the density of nearby and distant space? On a photographic plate nearby clusters are superposed on more distant ones, so that the images of their member galaxies are thoroughly scrambled. It is therefore extremely difficult, if not altogether impracticable, to separate the distant clusters from the nearer ones. However, it occurred to us that a statistical analysis of the plate as a whole might resolve the question. The analysis depends on the fact that any cluster at a great distance will look smaller and more tightly packed than one of the same size close by. On this basis we can picture roughly how a plate might look if it contained a disproportionate number of distant clusters (as required by the evolving model) or a constant density (steady state). Greatly exaggerated versions of these alternative distributions are pictured in the diagrams on page 111. Comparing them with Shane's plate (page 105), it is plain that one cannot decide on the actual distribution simply by glancing at his plot. But they indicate that a statistical analysis could decide which alternative better fits the actual picture.

Briefly, the method is as follows: Each map is divided into many small squares. The galaxies in each square are counted. If a square happens to cover part of a rich cluster, this square will contain substantially more than the average number of galaxies. In

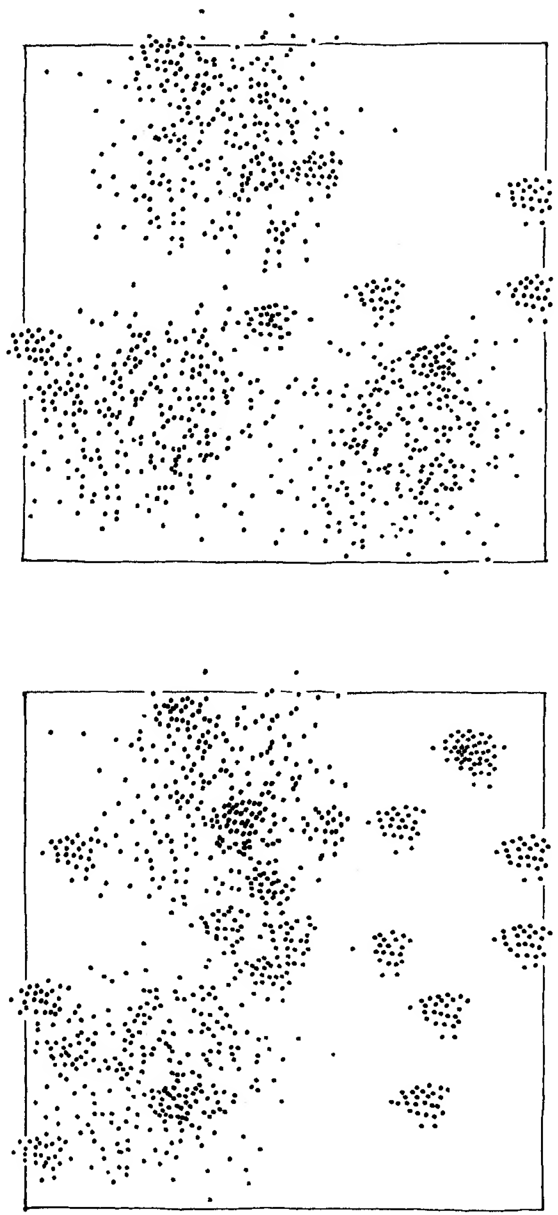


Evolutionary and steady-state universes give different values for density of clusters at increasing distances. For steady-state model (*left-hand sector*) the density is constant throughout all of space. For evolving model (*right-hand sector*) the density increases with distance. In this sector equal volumes contain from four clusters locally to seven at 900 million light years.

that case the adjoining squares are also likely to be at least partly within the cluster and to have a high galaxy count. So, too, the squares next to these and so on, but with decreasing probability of high counts as we get farther from the center of the cluster. The larger the cluster, the farther out the high count will extend. The plate is gone over square by square, and correlations are computed for adjacent squares, squares once removed, and so on. These correlations provide a basis for determining the distribution of cluster sizes in space and thus deciding between the opposing theories.

The idea is simple enough but it involves an enormous amount of computation, which is now in progress on a high-speed computer. At the moment we are analyzing plates made with the 20-inch telescope at the Lick Observatory. However, this instrument, with a range of only a few million light-years, does not cover a sufficient span of time to show detectable differences in density. Our hopes for a decision are tied to the new 120-inch telescope now nearing completion at the Lick Observatory. According to the present plan, as soon as this telescope is in operation a new survey of galaxies will be undertaken in co-ordination with the old survey. When these data become available for analysis, the crucial moment will have come. Preliminary computations indicate that the 120-inch telescope is powerful enough to penetrate far enough back in time to tell which of the two categories of cosmological theories is closer to reality.

Cluster plots made with large telescopes may settle the question of the expansion of the universe. An excess of small distant clusters over large near ones would indicate expansion. These diagrams are greatly exaggerated schematic illustrations of two possibilities. At the top is a plot which might represent a static universe. It contains three large clusters and ten small ones. At the bottom is a plot containing two large and twenty small clusters. A distribution of this type would be strong evidence in favor of the idea that the universe is expanding.



RADIO GALAXIES

by Martin Ryle

Just 25 years ago, when astronomers on Mount Wilson were beginning to get their first glimpse of the expanding universe and to measure the flight of distant galaxies, a radio engineer at the Bell Telephone Laboratories named Karl G. Jansky picked up some puzzling radio "static" which he decided must come from outside the atmosphere of the earth. It would have been impossible then to see any connection between the two events. But in the intervening 25 years, indeed mainly by rapid developments within the last 10, Jansky's static has ranged itself alongside the giant optical telescopes as a remarkable new window upon the universe. Radio astronomy gives us a totally new picture of our sun, of the Milky Way and even of what lies in interstellar space; but more than that, it now promises to extend our view into the depths of the universe and answer some of the central questions of cosmology.

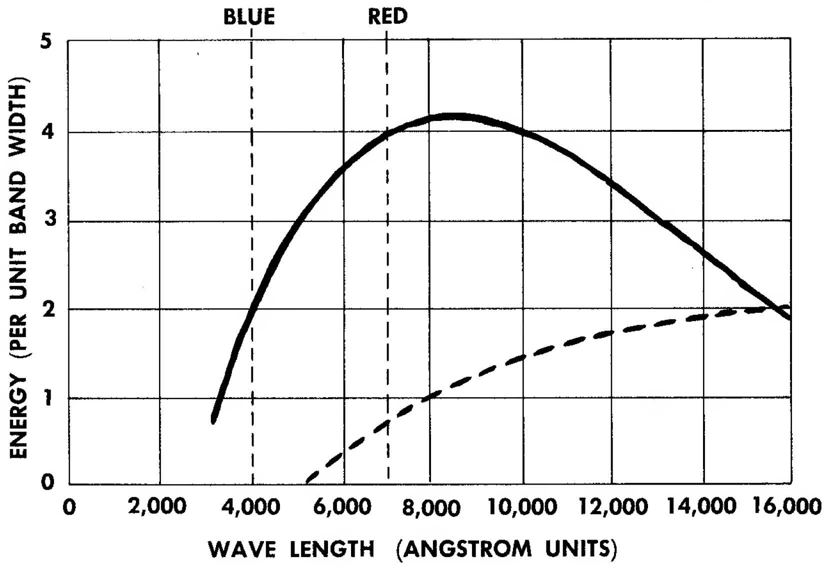
As Rudolph Minkowski relates in his chapter in this book (page 51), five years ago radio astronomers located a pair of colliding galaxies some 300 million light-years away. The radio signals from this collision, called Cygnus A, are sufficiently strong so that it could be detected even if it were at a much greater distance, beyond the range of the 200-inch telescope. Moreover, radio has an inherent advantage over light in probing to great distances. Reception is less weakened by the red shift. This shift is very substantial at great distances. At three billion light-years, for example, the light from galaxies (there moving away from us at half the speed of light) would be shifted so far toward the red that photographic plates could record only part of their spectrum in the visual range; the rest of their light would be lost

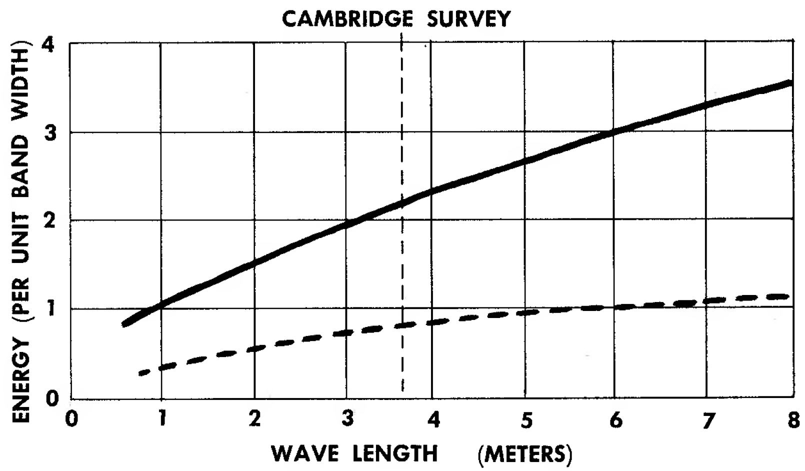
(see diagrams on page 114). Radio energy also suffers some loss with a wave length shift, but this loss is comparatively small, as analysis of the Cygnus A radio spectrum shows.

Apart from the advantage in detecting galaxies at greater distances, this difference between radio and light has other important consequences. Because of the increasing red shift, beyond a certain remote distance optical telescopes will fail to get any appreciable light at all, and the background of the sky will therefore appear dark. Radio telescopes, on the other hand, should receive the merged background radiation to much greater distances, some of it originating from extremely distant sources. Thus radio astronomy offers the possibility of sampling the material content of the universe in very remote space and thereby testing cosmological theories.

How has this possibility come about? To see how Jansky's discovery has become a tool for exploring the universe we must briefly examine the nature of the technique and the significant developments that stimulated interest in it. At first, radio astronomers could find only diffuse regions of radio emission in the sky. The basic problem of radio astronomy is that radio waves are so much longer than light waves. To equal the resolving power of the human eye would require a radio telescope about 10 miles wide. A second major problem is the weakness of the celestial radio signals as we receive them on the earth: the faintest now detected have only a hundred-millionth of the power of a television signal. There is a limit to the sensitivity that can be achieved in a receiver.

The cure for both problems, of course, is larger antennas. Increasing the size of the antenna improves resolution and collects more radio energy from the source, just as a larger mirror in an optical telescope collects more light from a star. The wartime radar "dishes" first used gave way to larger and larger bowls and then to arrays of linked antennas spread over acres of countryside. As antennas grew in size, radio astronomers were





able to narrow down the radio sources to smaller and smaller regions.

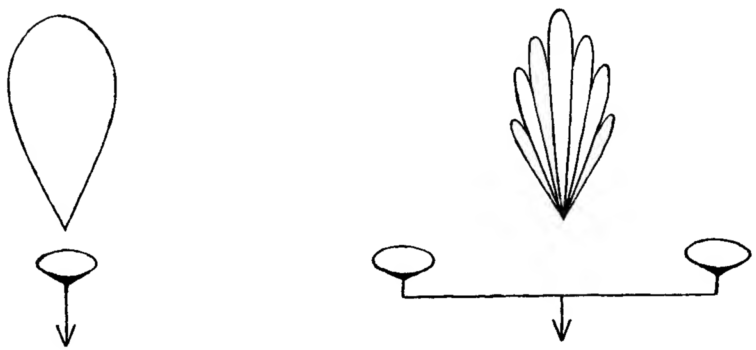
It was about ten years ago that the first so-called "radio stars" were discovered. They were located as narrow centers of radio emission—almost "point" sources of radio energy. Quite evidently these objects were not ordinary stars: none corresponded to a visible star, but a few coincided with nebulae of peculiar characteristics. One of the first objects pinpointed as a radio star was the well-known Crab Nebula, an expanding gaseous cloud in our galaxy which represents the remains of a supernova whose explosion was recorded by Chinese astronomers in the year 1054.

The finding of radio stars naturally stimulated efforts all over the world to erect radio telescopes of higher resolving power. Their development has taken two main lines. The first followed the lead of optical astronomy in building bigger and bigger instruments of the reflector type—taking the form in the radio case of a huge paraboloid bowl. This approach is illustrated by the 50-foot dish at the Naval Research Laboratory in Washington, the 75-foot one in the Netherlands, and finally the 250-foot giant at the University of Manchester in England. The paraboloid type

Effect of red shift on light waves and radio waves is contrasted in these two charts. The solid curve in chart at top shows the distribution of energy across the optical spectrum, from blue to red, in the light from a nearby star. The broken line in that chart shows how the red shift of a distant galaxy may cut off reception of light in the blue region to which photographic plates are most sensitive. The chart below similarly shows the distribution of energy in the spectrum of radio waves received from a nearby and a distant source. Though the radio spectrum is shifted, the reduction in the intensity of the radio energy received is slight. The Cambridge radio-source survey was made at a wave length of 3.7 meters.

of antenna has many advantages, especially maneuverability in scanning the sky. For the study of radio stars, however, it is necessary to tune in to the longer radio wave lengths, where the stars' emission is strongest. To obtain sufficiently fine resolution at these wave lengths any feasible dish is too small: larger antennas are needed. Most of the radio star observations so far have been made with extended antenna systems based on the interferometer principle.

The simplest of these systems uses two dishes stationed a considerable distance apart. Where a single dish, scanning the sky,

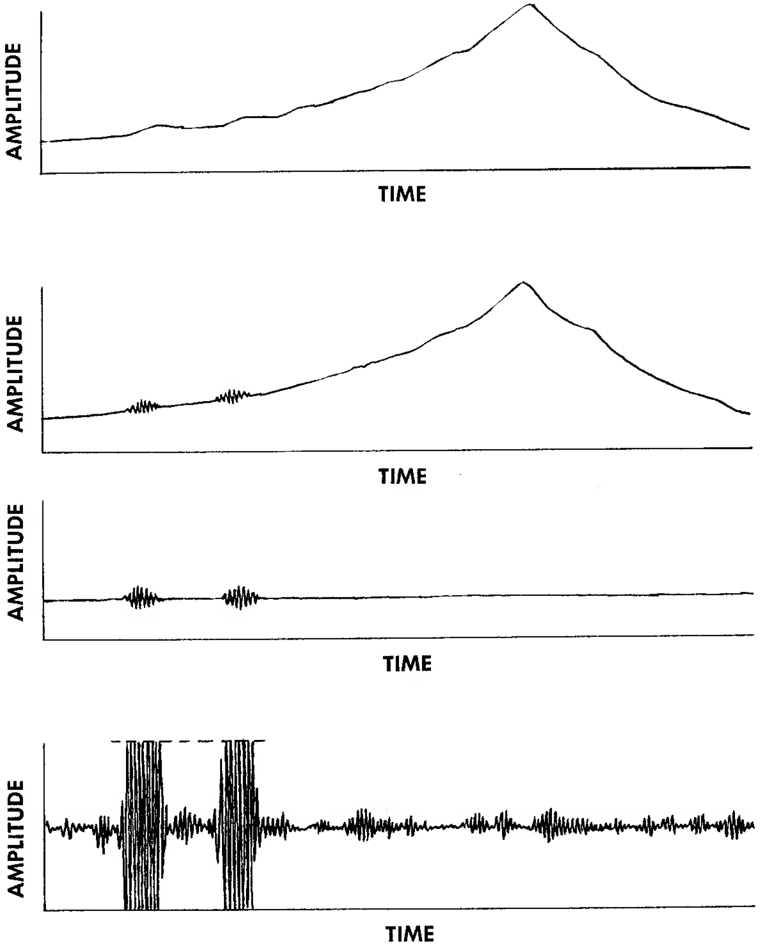


Reception patterns of a single-antenna telescope and an interferometer are contrasted. The large single lobe at left indicates the wide angle of resolution of the single antenna. The interferometer pattern, represented by multiple lobes, achieves finer resolution.

records the aggregate radiation from a region, the two-dish system picks out radio stars from the background (see charts opposite). With a device called phase switching, the background radiation can be wiped out so that the receiver records only the "point" sources. By increasing the sensitivity of the receiver, weaker radio stars can be brought into the picture, as the last chart in this group shows. In this case the "images" overlap, however, and fainter sources can be detected only by improving the resolution. One way of doing this is to use two narrow antennas oriented in different directions, one receiving a pattern in the north-south plane and the other in the east-west; it is thereby possible to resolve sources at the point where the reception patterns coincide. B. Y. Mills in Australia has recently completed a large system of a similar type in which the two narrow antennas cross each other (the "Mills cross"). This system corresponds to a single dish of high resolving power (see diagrams on page 120).

By 1950 our group at the University of Cambridge had located 50 radio stars in the northern hemisphere of the sky, and workers in Australia had found a similar number in the southern hemisphere. Of all these objects, only a few could be identified with luminous bodies. One was the Crab Nebula; another was the galaxy known as M 87, which has an unusual bright "jet" near its nucleus. The light from both the Crab Nebula and M 87 has been found to be strongly polarized, which may be a clue to the origin of the radio emission. It seems probable that the radio energy and part of the light from these two sources is generated by the motion of high-energy electrons in a magnetic field.

The immediate question was: Where did the unidentifiable radio stars lie in space? Were most of them within our own galaxy, or did many of them, like M 87, represent objects outside our galaxy? For information on this question our group made a special study of the two strongest radio stars so far located—one in the constellation of Cassiopeia and the other in Cygnus. It

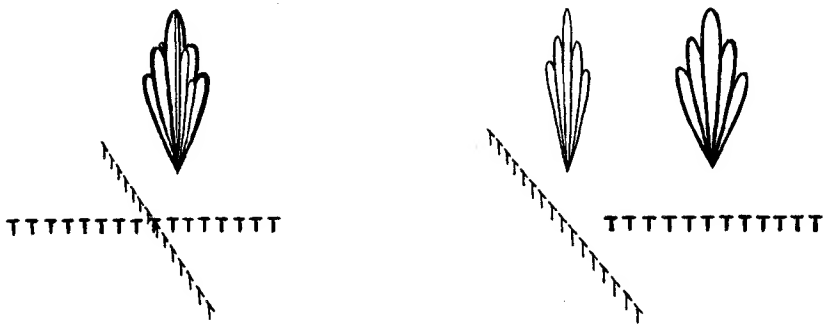


Amplitude curves produced by various radio telescope systems are diagramed here. In the curve at the top, recorded by a single antenna, it is difficult to distinguish the radio stars from the smooth variation in amplitude across the Milky Way. The second curve, recorded by an interferometer, registers the existence of two discrete sources, along with the variation in background. In the third curve, the background variation is suppressed and only the discrete sources are recorded; this is achieved by "phase-switching." With higher sensitivity, a phase-switching interferometer will produce the curve at bottom, which shows not only the strong sources but also numerous weak sources, so close together that they overlap.

soon became clear that to pinpoint the sources with sufficient accuracy to identify them with any visible object would require higher resolving power than we had available. A new radio telescope of the interferometer type was therefore built for the special purpose of identifying these two sources. During the spring of 1951 F. G. Smith succeeded in narrowing down the positions of both radio stars to small areas about one hundredth the size of the earlier locations.

Walter Baade then photographed the positions with the 200-inch telescope on Palomar Mountain. The Cassiopeia source turned out to be a gaseous nebulosity in our own galaxy, perhaps the remains of a supernova. But the source in Cygnus was identified with a faint nebula whose red shift showed it to be about 300 million light-years from us—far outside our galaxy. The radio emission of this object was calculated to be about equal in energy to its light emission. Further study, as described in Minkowski's chapter, revealed that the object, named Cygnus A, was actually a pair of colliding galaxies.

The discovery was at once recognized to be of profound importance for cosmological research. Not only did it open the prospect of extending our view beyond the reach of the 200-inch telescope but it promised specific information of a kind that could



Orientation of antennas at right angles increases the resolution of radio interferometers. Australian "Mills cross" is diagramed at left; another Cambridge installation, at right. The two systems register only sources that are picked up by both of their antennas.

never be obtained by optical means, however large the telescope. Extragalactic radio sources as intense as Cygnus A must be extremely rare: it is so much "brighter" than any other detected radio star (except for identified objects in our galaxy) that there cannot be more than one such source in every 100 million normal galaxies. Objects of this kind should therefore be detectable even at very great distances. If a way is found to recognize colliding galaxies and to determine what fraction of the radio stars are of this type, it should become possible, by statistical analysis, to make direct deductions about the most distant parts of the universe.

With the object of detecting more radio stars to increase the number for analysis, and of obtaining more accurate positions, a new, considerably larger telescope was constructed at Cambridge in 1952. It is a double interferometer consisting of four aerials each 320 feet long (see diagram on page 123). The aerials

can be tilted, and with the earth's rotations they can survey the whole sky at Cambridge.

A comprehensive survey with this instrument, completed early in 1955, located 1,936 radio stars. The first question to be determined, of course, is whether they lie in our own galaxy or outside it. Thirty of them give indications of being part of our system: their diameter is comparatively large; they tend to be concentrated near the plane of the Milky Way, and several have been identified with gaseous nebulosities within the galaxy.

The remaining 1,906 are "point" sources, distributed uniformly across the sky. Very few of them can be identified with visible objects, and these give us no enlightenment on the population as a whole. Two are near supernovae in our system (discovered respectively by Tycho Brahe in 1572 and by Johannes Kepler in 1604); but, on the other hand, some coincide with extragalactic objects, including one in the Perseus cluster which has been identified as a collision between galaxies.

Even intensive inspection of photographic plates, then, fails to answer the question whether most of the radio stars are members of our galaxy or outside it. Their uniform distribution across the sky might suggest that they are extragalactic, because galaxies are distributed uniformly over the whole sky while our own system is concentrated near the circle of the Milky Way. But it is conceivable that we are surrounded by radio-emitting objects within our galaxy—a class of objects which we cannot see and whose existence we have not hitherto suspected. Since we do not know what the intrinsic intensity of these radio sources is, we cannot tell how far away they are—whether close to us or outside the galaxy.

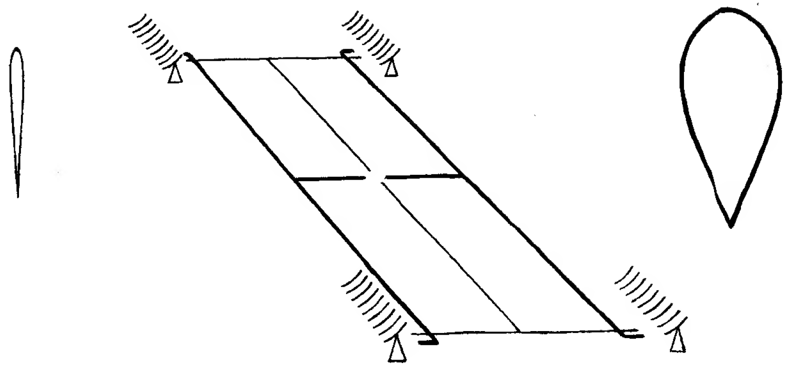
We undertook to attack the problem in another way: an analysis to determine whether the density of distribution of the radio stars changes with distance from us. The basis of this analysis is the law that the intensity of a source of radiation (light or radio) decreases as the square of the distance. As a con-

sequence of the law, it is possible to calculate how many stars within a spherical region will have more than a given brightness, if the stars are uniformly distributed through it. For example, if we count a certain number of stars of one brightness, we should find that the number of stars with at least one quarter of that brightness is eight times as great. But the ratio will vary from this if (1) the average density of the sources changes with distance from us, or (2) the average power of the sources changes with distance.

When this method of analysis was applied to the 1,906 radio stars, it was found that there were too many weak sources in relation to the intense ones. This result implied that one or the other of the nonuniform alternatives was true: that is, either the density of sources or the average power emitted increases with distance. Careful checks of both the data and the interpretation, including an analysis by an independent method, confirmed this conclusion.

Now, the result may be interpreted in either of two ways. We may suppose that the radio stars are within the galaxy but we happen to lie in a region where their density is abnormally low, so that the density rises with distance from us. But then it is very difficult to explain why the radio stars are distributed symmetrically over the whole sky, for the supposition of a variation in density implies that these objects are not spread uniformly over the galaxy. Moreover, the aggregate amount of radiation from these radio stars is so large that they could not be a small eddy: they would have to occupy a considerable part of the galaxy.

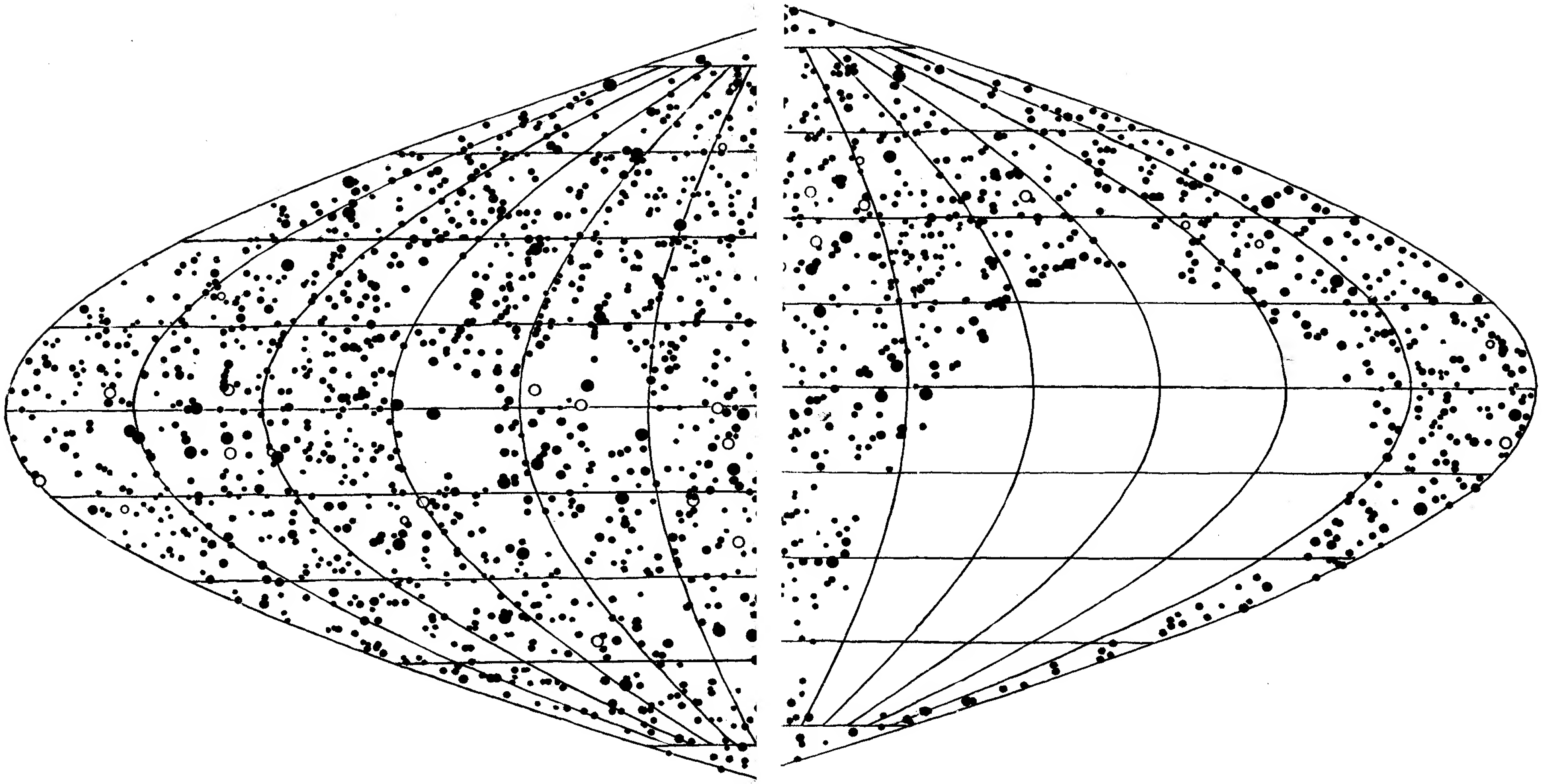
If, on the other hand, we suppose that most of the radio stars are outside our galaxy, we can see a consistent and reasonable picture. Calculations of the aggregate amount of radiation from the radio stars led our group to conclude that the region containing the detected radio stars must extend out to at least 500 million light-years, and that a large fraction of the stars are extragalactic sources of great power—comparable to that of Cygnus A.



The Cambridge interferometer is diagramed here. The circuit indicated by heavy lines sets up a high-resolution reception pattern (*left*). The circuit indicated by the light lines produces a pattern with lower resolution (*right*).

The apparent increase in density of the sources with distance may then be accounted for as an effect associated with the large red shifts of the distant sources. Further, it becomes possible to explain why no visible objects appear at the positions of most of the radio stars: if these stars are colliding galaxies, only a few dozens of them lie within reach of the 200-inch telescope; the rest are beyond the visible range.

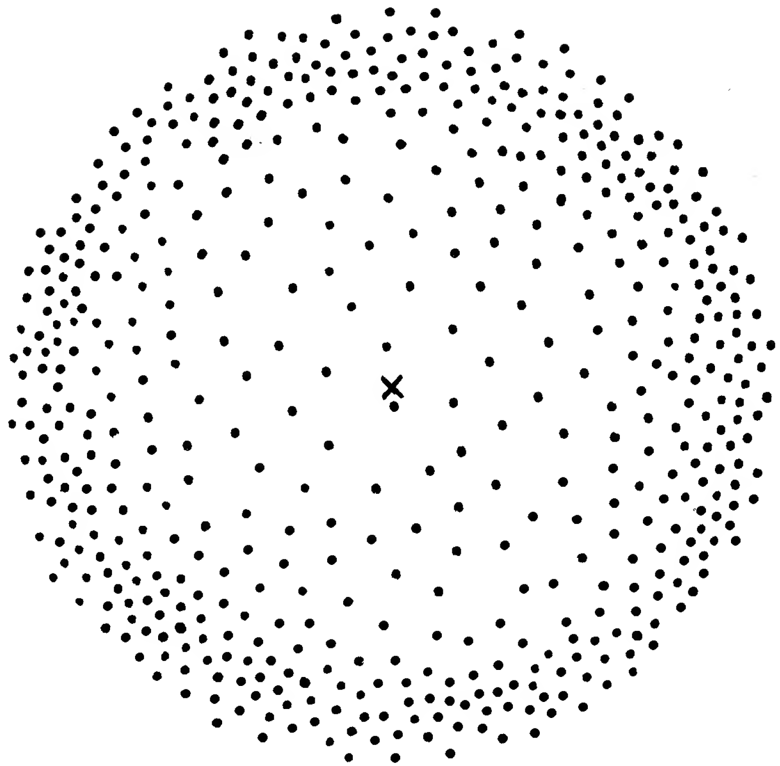
The Cambridge conclusion about the distribution of radio stars in space far beyond our galaxy has been questioned by workers in Australia. A survey with the Mills-cross radio telescope has failed to show a marked excess of faint sources such as was found by the Cambridge group. The Australian survey, however, has not yet covered a large area of the sky, and it does indicate that radio stars are not distributed uniformly with distance. At Ohio State University observations with a pencil-beam radio telescope by the radio astronomer John D. Kraus have confirmed the excess of faint sources. The question may be settled



Radio sources located by the University of Cambridge survey are shown in this sky map. The map is drawn on galactic coordinates; *i.e.*, the north pole of the galaxy is at the top, and the plane of the Milky Way is the equator. Open circles mark sources of large diameter; the intensity of the other sources is indicated by the size of the black dots. The vacant space at the right is the portion of the sky that cannot be seen from Cambridge.

conclusively within the near future by completion of the Australian survey and of a new survey, using higher resolving power, which is now under way at Cambridge.

If these surveys verify that the density of radio sources in space does indeed increase with distance, they should help to make possible a decision between the evolutionary and steady-state theories of the universe. If most of the radio stars are in fact collisions between galaxies, such encounters apparently are considerably more frequent in distant space (perhaps billions of light-years away) than near us. This disparity would argue against the steady-state hypothesis that the density of matter in space remains constant. The radio signals we are now receiving from distant collisions started on their way billions of years ago. If the evolutionary theory is correct, the universe should have been denser then, and encounters between galaxies more likely. Thus our present conclusions from the radio work at Cambridge support the evolutionary view.



Distribution of radio sources in space, as deduced by the group at Cambridge, is diagramed. The distribution of sources is symmetrical, but from the number of sources of different intensity it appears that they must increase in density with distance.

PART 6 CRITIQUE

COSMOLOGY AND SCIENCE

by Herbert Dingle

Herbert Dingle has been a firsthand observer of some of the leading developments in twentieth-century science. He is professor of the history and philosophy of science at University College, London. He was born in 1890 and completed his education at the Imperial College of Science and Technology. "I began my scientific career under the late Professor A. Fowler at a time when his spectroscopic work was essential to the progress of the new Bohr theory of the atom. I was thus admitted to the actual scene of development of the quantum theory. At about the same time Einstein's general relativity theory came to the fore through the 1918 eclipse, and I had the privilege of working with A. N. Whitehead, who read the manuscript of my first book on the subject; and of being in close touch with A. S. Eddington at the Royal Astronomical Society, so that I was at close quarters with two different and important views of the philosophical aspects of the theory. Add to this a natural taste for the broader aspects of scientific philosophy, and my type of career is pretty well determined." Dingle has been a leading figure in British astronomy; from 1951 to 1953 he was president of the Royal Astronomical Society.

COSMOLOGY AND SCIENCE

by Herbert Dingle

SINCE the advent of Einstein's general theory of relativity the subject of cosmology has assumed a much more prominent position in the world of physical science than it had occupied in recent times. It is natural to assume, as many do, that this is a result of the development of our knowledge and our means of observation—our ability to examine very distant regions of space with powerful telescopes, and the growing ability of mathematicians to make trustworthy deductions from the seen to the unseen. In the earlier stages of science, so it is thought, men had to confine their consideration to the parts of the universe that were within reach, and so were concerned not with the universe as a whole but only with the local and the partial.

Nothing could be further from the truth. The fact is that during by far the greater part of its history astronomy has been nothing but cosmology. Until the seventeenth century the one aim of the astronomer (apart, of course, from such practical applications as navigation, time measurement, astrology and so on) was to describe the working of the whole system of the universe as he saw it. But early in the seventeenth century two independent things happened at about the same time: the collapse of Aristotelian cosmology and the birth of the telescope. Thenceforward observational astronomy advanced step by step and cosmology lay dead. Before very long a promise of a resurrection of cosmology in a new form came with the conception of universal law introduced by Newton's great work. But universal law told you nothing about the actual structure of the universe. It told you how any possible universe would operate, and therefore could not distinguish between one possible universe and another.

The distinctive feature of our time, underlying the present revival of cosmology, is that our improved knowledge of universal law, together with our extended knowledge of the actual content of the celestial spaces and the behavior of the bodies in them, have reached a point at which they can be brought together into a single scheme. The freedom allowed by the former can be so limited by the latter as to give us a first approximation to an understanding of the whole universe which is capable of satisfying both our reason and our observation of the heavens.

It is the purpose of this article to compare the points of view of the ancient and the modern cosmologies and to indicate how the one gave place to the other. This is not merely of historical interest. It can serve a very practical purpose in enabling us to avoid the errors that brought the work of our predecessors to grief.

To appreciate what kind of thinking created the cosmology of the Greek pioneers, beginning, as it inevitably had to in early times, from the natural assumption that the earth was the center of the universe, we must understand a fundamental characteristic of Greek thought—which is at variance with the scientific outlook. They presupposed certain *principles*, which were assumed to be inviolable and were accepted without question. If appearances seemed to contradict them, then the appearances were deceptive. These principles must not be confused with what we call rational necessities, such as, for instance, the axiom that things which are equal to the same thing are equal to each other. This cannot conceivably be violated within the framework of ordinary geometry, because it is inherent in the definition of equality, and not an assertion about the characteristics of things. But the ancient Greek principles (of which those that persisted longest were due mainly to Aristotle) were assertions about the characteristics of things. For example, they asserted that the only activity possible to heavenly bodies was perfectly uniform and circular movement, and that apart from such eternal circulations no change of

any kind could take place in the heavens. If a sunspot or some other change appeared to occur on the face of the sun, it could be taken only as an appearance and must be due to something in the earth's atmosphere passing between the observer and the sun. And similarly, since the planets appeared not to move in circles at uniform speed, the apparently erratic movements of each planet must be the resultant of a set of circular movements.

The aim of Greek cosmology was to arrive at the complex system of interlocking spheres in motion that made up the universe. The individual heavenly bodies themselves were merely straws from which to determine how the wind blew: the wind was the important thing. Geometers of genius such as the Greeks produced were able to represent the observed movements of the planets with an accuracy equal to that of their imperfect observations at any given time, but as time went on the discrepancies between the geometrical requirements and the observed positions of planets increased, and so more spheres were introduced to annul them. This went on throughout the Middle Ages, until by the sixteenth century more than 80 spheres were necessary to account for the observed movements, and even that number did it very imperfectly.

Now, if astronomy had stood by itself as an isolated study, this complexity might well have stimulated efforts at reform earlier than it did. But the other spheres of study were so thoroughly interwoven with astronomy that it was impossible to reform astronomy without also reforming physics, chemistry, physiology, psychology and theology (to use modern terms for subjects not so clearly differentiated from one another then as now). It could not be done without upsetting the whole scheme of belief. The universe was then a universe in a much more literal sense than it has ever been since. A diagram taken from a textbook of the time makes this very evident: Heaven, where God dwelt with the elect, had a location which was as much a part of the physical universe as the earth and the cosmic spheres. Each of the planets had its particular influence on human tem-

perament: thus we get our adjectives mercurial, martial, jovial, saturnine. A human calamity was a *dis-aster*—against the stars. An unnatural action was *ex-orbitant*—out of orbit. Terrestrial bodies were compounded of four elements—earth, water, air and fire—each of which tended to seek "its own place," and the heavenly bodies were composed of a perfect, unchangeable fifth element—a "quintessence"—which had no parallel on the changeable earth. And so on.

Into this closely interwoven scheme it was clearly very hazardous to introduce any modification of a single part, because of its possible unforeseen effect on the whole. Nevertheless, by the sixteenth century the cosmic machinery of spheres had become so unwieldy that Copernicus, a man dominated by the mathematician's passion for simple generalization, ventured to make what seemed to him the very slight change of transferring the center of the universe from the earth to the sun. By this device he was able to reduce the number of cosmic spheres by more than half.

He made no other change, nor did he realize that any other was necessary. He clung as firmly as the most orthodox medieval philosopher to the machinery of spheres and to the Aristotelian principles of perfect celestial substances and uniform circular motions. He thought he could simplify without destroying. But in fact this one small change shattered the whole medieval universe. By the time of Galileo, some three quarters of a century later, it had become clear that there was no need for any spheres at all, and the simple change that we would now describe as no more than choosing a different origin of co-ordinates had generated a conflict of world views such as the world had never before known. The cosmic conception of more than 2,000 years' standing had in fact received its deathblow.

Tycho Brahe and Johannes Kepler completed the job—the former by observing change in the supposedly immutable sphere of the stars and the latter by showing that the motion of a planet could be represented by a simple ellipse. Galileo and the invention of the telescope gave birth to a new astronomy—and to a new philosophy of science. Soon afterward the work of Newton opened up the possibility of a kind of cosmology not previously conceived.

The contrast between the old and the new attitude can be fittingly introduced by the words of Newton himself: "To tell us that every Species of Things is endow'd with an occult specifick Quality by which it acts and produces manifest Effects, is to tell us nothing: But to derive two or three general Principles of Motion from Phaenomena, and afterwards to tell us how the Properties and Actions of all corporeal Things follow from those manifest Principles, would be a very great step in Philosophy, though the Causes of those Principles were not yet discover'd." What Newton referred to as "occult qualities" were neatly illustrated by the notions of "gravity" and "levity" in the old cosmology: this body fell downward because it was subject to gravity, that body rose upward because it was subject to levity. The principles were conceived merely because they seemed fitting. Neither gravity nor levity had any characteristics by which it could be identified except the movement it was supposed to explain. They were mere names for the phenomena observed, masquerading as causes of those phenomena. The whole scheme of cosmological principles resolved itself into a series of tautologies. It was logically impeccable but scientifically barren—completely unproductive of knowledge.

The Galilean-Newtonian philosophy, on the contrary, brought knowledge in apparently boundless measure but was logically outrageous. From a few phenomena, or experiments, it proposed to derive principles to be applied universally. Some bodies attract one another; this is a body; therefore it attracts every other body in the universe. No more patently invalid syllogism could be imagined; it was an error in Aristotelian logic of which even the youngest scholastic child could hardly have been capable.

But it worked. And not only so, but similar generalizations

later in other fields were found to work, and they go on working. We have never known heat to flow by itself from a colder to a hotter body, and we take it that it never has nor ever will anywhere in space. The brightness of a lamp in our laboratory falls off as the square of the distance as we walk away from it; hence we infer what the brightness of a distant galaxy must be. That is science. Its assertions about the universe are unlimited generalizations from a few momentary observations at a point in space.

From a purely logical point of view scientific cosmology would appear to have no justification, to be a gigantic impertinence. It is saved from this by a frank recognition by scientists of what it is and what its limitations are. The work of three centuries has shown that the scientific approach is on unassailable ground when it declares itself to be the best prescription yet devised for obtaining knowledge of the relations between phenomena. Whether or not its generalizations have any right to be regarded as the *truth*, they lead to further knowledge—which, so far as we can see, would be quite unobtainable otherwise. But they do this only on the condition that we abandon them the moment we see that they cease to hold. They originate in phenomena and they are at the mercy of phenomena.

The Aristotelian general principles, on the other hand, were conceived a priori, independently of phenomena, and phenomena were distorted at liberty so as to exemplify them. The problem was to "save the phenomena." The basic principles themselves could not be threatened; it was the phenomena that stood in need of salvation.

Contrast with this the point of view that caused modern cosmologists to discard the Newtonian cosmology for the totally different Einsteinian cosmology—on the basis of the tiniest of differences between theory and observation. At the behest of measurements so fine that only in this age have they become possible at all, we have thrown over the whole picture of a universe in which bodies are pushed along by a conspiracy of alien forces and have substituted in its place a system of free bodies moving

along the paths which, so to speak, offer them the easiest course. We are prepared to revise our principles to fit observation because man's whole experience in seeking knowledge has taught us that nature is far more likely to follow in the large the laws we observe her to follow in the small than to behave according to our intuitive ideas of decorum.

In most branches of science this philosophy has been adhered to unswervingly. The activities of experimental physicists, chemists and biologists are for the most part beyond reproach: we learn where we can, generalize from observation and extend the generalizations until they are found to fail, when we replace them by wider ones. But in cosmology this restraint does not always hold. Some cosmologists have returned to the discredited practice of inventing arbitrary general principles, with no justification except that they seem "right," and fitting phenomena to the requirements of the principles. We have been presented, for example, with a "cosmological principle" which demands that, on the large scale though not on the small (where it might readily be tested), every part of the universe must be exactly the same as any other. This has been extended to a "perfect cosmological principle" which says that the uniformity of the universe must *always* hold true, its general appearance being eternally the same. If the scientific procedure of generalizing from observation leads to anything inconsistent with this, it must be wrong. The principles are of necessity inviolable, and all phenomena must be interpreted in accordance with them. There could scarcely be a more complete return to a kind of philosophy which we once thought had been abandoned forever.

Anyone acquainted with the history of cosmology can recognize the cosmological principle and the perfect cosmological principle as having precisely the same nature as perfectly circular orbits and immutable heavens. Indeed, the perfect cosmological principle is largely identical with the Aristotelian principle of unchanging celestial regions.

We should not confuse the cosmological principle with the valid scientific procedure of assuming, for purposes of investigation, that the rough approximation to homogeneity which we observe in space as far as we can see it extends to the whole of space. That is the normal type of scientific generalization, and it is usually made in relativistic cosmology in order to restrict the almost limitless field of investigation that would be open if all possibilities were taken into account. We calculate what it demands of the yet unexamined regions and hold ourselves ready to discard it if they should fail to meet those demands. The cosmological principle, however, alters observed facts to make them accord with its requirements. Observation appears to show that the density of matter in the universe is continually decreasing. This cannot be so, says the principle: unobservable matter must be in process of creation out of nothing in just the right amount to keep the density constant. We have no evidence of such matter. The only reason for supposing its creation is that otherwise the perfect cosmological principle would be violated, and the only reason for supposing that this cannot happen is that a few mathematicians would not like it. It seems an insufficient reason.

Such atavisms notwithstanding, in the main the cosmologies of our day are founded upon the scientific procedure: generalizing from what we know and then testing our generalizations by observations on a larger scale. We have reached a point at which the theoretical generalizations can be compared with a sufficiently wide field of observation to entitle us to think that we can truly say something about the universe as a whole—not only about its laws of operation but also about the particular arrangement of bodies that exhibit those laws. Doubtless our present views will be modified and enlarged in the future, but they do embrace a far wider range of knowledge than has ever been possible before. What that knowledge comes to as of A.D. 1956 is well summarized in this book.

BIBLIOGRAPHY

READERS interested in further reading on the topics covered in this book may find the list below helpful. It is *not* a bibliography of source material. The books chosen are for the most part addressed to the general reader; they include also some of the more accessible textbooks and survey volumes. The list is by no means exhaustive. Nor does it embrace the full range of interest in this book, since much of the work reported here is not yet represented in the pages of any other book. (All these chapters appeared originally in the September, 1956, issue of SCIENTIFIC AMERICAN.)

THE UNIVERSE

Frontiers of Astronomy. Fred Hoyle. Harper & Brothers, 1955.

A History of Astronomy from Thales to Kepler. J. L. E. Dreyer. Dover Publications, Inc., 1953.

The New Astronomy, a SCIENTIFIC AMERICAN book. Simon and Schuster, 1955.

THE ORIGIN OF THE ELEMENTS

"Element-Building Reactions in Stars." William A. Fowler and Jesse L. Greenstein in *Proceedings of the National Academy of Sciences*, Vol. 42, No. 4, pages 173–180; April, 1956.

"Stellar Evolution and the Synthesis of the Elements." W. A. Fowler, G. R. Burbidge and E. Margaret Burbidge in *The Astrophysical Journal*, Vol. 122, No. 2, pages 271–285; September, 1955.

"Theory of the Origin and Relative Abundance Distribution of the Elements." Ralph A. Alpher and Robert C. Herman in *Reviews of Modern Physics*, Vol. 22, No. 2, pages 153–212; April, 1950.

THE CONTENT OF GALAXIES

"A Larger and Older Universe." George W. Gray in SCIENTIFIC AMERICAN, Vol. 188, No. 6, pages 56–66; June, 1953.

THE EVOLUTION OF GALAXIES

"Astronomy—the Birth of Stars from Interstellar Clouds." Lyman Spitzer, Jr., in *Journal of the Washington Academy of Sciences,* Vol. 41, No. 10, pages 309–318; October, 1951.

Galaxies. Harlow Shapley. The Blakiston Company, 1943.

The Realm of the Nebulae. Edwin Hubble. Yale University Press, 1936.

"A Receiver for the Radio Waves from Interstellar Hydrogen: 1. The Investigation of the Hydrogen Radiation." C. A. Muller in *Philips Technical Review,* Vol. 17, No. 11, pages 305–315; May, 1956.

COLLIDING GALAXIES

"Identification of the Radio Sources in Cassiopeia, Cygnus A, and Puppis A." W. Baade and R. Minkowski in *The Astrophysical Journal,* Vol. 119, No. 1, pages 206–231; January, 1954.

"Stellar Populations and Collisions of Galaxies." Lyman Spitzer and Walter Baade in *The Astrophysical Journal,* Vol. 113, No. 2, pages 413–418; March, 1951.

THE EVOLUTIONARY UNIVERSE

The Creation of the Universe. George Gamow. The Viking Press, 1952.

"Modern Cosmology." George Gamow in SCIENTIFIC AMERICAN, Vol. 190, No. 3, pages 54–63; March, 1954.

The Primeval Atom. Georges Lemaître. D. Van Nostrand Company, Inc., 1950.

THE STEADY-STATE UNIVERSE

Cosmology. H. Bondi. Cambridge University Press, 1952.

"The Steady-State Theory of the Expanding Universe." H. Bondi and T. Gold in *Monthly Notices of the Royal Astronomical Society,* Vol. 108, No. 3, pages 252–270; 1948.

THE RED SHIFT

The Expansion of the Universe. Paul Couderc. Faber & Faber, Ltd., 1952.

The Realm of the Nebulae. Edwin Hubble. Yale University Press, 1936.

"Red-Shifts and Magnitudes of Extragalactic Nebulae." M. L. Humason, N. U. Mayall and A. R. Sandage in *The Astronomical Journal,* Vol. 61, No. 3, pages 97–162; April, 1956.

THE DISTRIBUTION OF GALAXIES

"Comparison of the Synthetic and Actual Distribution of Galaxies on a Photographic Plate." E. L. Scott, C. D. Shane and M. D. Swanson in *The Astrophysical Journal,* Vol. 119, No. 1, pages 91–112; January, 1954.

"Frequency of Separation and Interlocking of Clusters of Galaxies." Jerzy Neyman and Elizabeth L. Scott in *Proceedings of the National Academy of Sciences,* Vol. 39, No. 8, pages 737–743; August, 1953.

"On the Problem of Expansion of Clusters of Galaxies." J. Neyman and E. L. Scott in *Studies in Mathematics and Mechanics Presented to Richard von Mises*; Academic Press Inc., 1954.

"On the Spatial Distribution of Galaxies: A Specific Model." J. Neyman, E. L. Scott and C. D. Shane in *The Astrophysical Journal,* Vol. 117, No. 1, pages 92–133; January, 1953.

"Spatial Distribution of Galaxies—Analysis of the Theory of Fluctuations." Jerzy Neyman and Elizabeth L. Scott in *Proceedings of the National Academy of Sciences,* Vol. 40, No. 10, pages 873–881; October, 1954.

RADIO GALAXIES

"A Detailed Radio Map of the Sky." J. D. Kraus and H. C. Ko in *Nature,* Vol. 175, No. 4,447, pages 159–161; January 22, 1955.

"The Isotropic Component of Cosmic Radio-Frequency Radiation." J. R. Shakeshaft in *The Philosophical Magazine,* Vol. 45, No. 370, pages 1,136–1,144; November, 1954.

"The Observation and Interpretation of Radio Emission from Some Bright Galaxies." B. Y. Mills in *Australian Journal of Physics,* Vol. 8, No. 3, pages 368–389; September, 1955.

Radio Astronomy. J. L. Pawsey and R. N. Bracewell. Oxford University Press, 1955.

"Radio Stars and Their Cosmological Significance." Martin Ryle in *The Observatory,* Vol. 75, No. 887, pages 137–147; August, 1955.

"Spatial Distribution and the Nature of Radio Stars." M. Ryle and P. A. G. Scheuer in *Proceedings of the Royal Society of London,* Series A, Vol. 230, No. 1,182, pages 448–462; July, 1955.

COSMOLOGY AND SCIENCE

Dialogue on the Great World Systems in the Salusbury Translation. Edited by Giorgio de Santillana. University of Chicago Press, 1953.

Opticks: or A Treatise of the Reflections, Refractions, Inflections & Colours of Light. Sir Isaac Newton. G. Bell & Sons, Ltd., 1931.

"Science and Modern Cosmology." Herbert Dingle in *Science,* Vol. 120, No. 3,118, pages 513–521; October, 1954.

The Scientific Adventure: Essays in the History and Philosophy of Science. Herbert Dingle. Sir I. Pitman & Sons, 1952.